DIGITAL LIBRARIES
University Libraries

UNT
EST. 1890

Improving Subjects in the Digital Collections with Data

White Paper

Researchers: Hannah Tarver | Digital Projects Unit
Chassidy Miles | User Interfaces Unit
Rachael Zipperer | Special Collections

December 2021

Acknowledgements

Table of Contents

1. Introduction

One key function of libraries has long been to connect users to information. Historically, libraries documented information about resources in card catalogs indexed according to three access points: title, author, and subject. Despite the relative difficulty and time required to assign subjects, librarians understand the importance of helping users find materials based on a particular topic of interest. Additionally, helping researchers or other users find multiple resources with related content relies on consistent use of subject terms and controlled headings (United States Office of Education, 1930).

As catalog records moved online, it became less expensive to add more subjects to individual records. Similarly, some online catalogs and most digital libraries have relaxed subject standards to allow for multiple controlled vocabularies, uncontrolled keywords, or combinations of subject usage. For printed text items, OCR[1] files allow full-text searching of content; however, for items with handwritten or image content, the only access point for users to find relevant materials is based on information included in the metadata records. As digital libraries add more historic photographs to their online holdings, this reliance on subject metadata reflects a desire to make these and other cultural heritage materials findable to a wider audience than local users of a library or archive.

In fact, subjects have remained a core component in the shift to online metadata records. A 2019 survey of digital libraries found that across 142 repositories, the `subject` field was the sixth-most-required field (in 36 of the repositories). Along with `date`, `subject` was also the second-most-included field -- either required, recommended, or optional in 84.5% (120) of the repositories (Gentry et al., 2020). Further response-level analysis has determined that `subject` is actually the most-frequently- included field in 123 repositories, adjusted for 3 responses that did not designate whether `subject` is required, recommended, or optional in those repositories.[2] Additionally, many digital libraries have multiple subject-based fields such as `genre` or `keyword`.

Not all digital libraries handle subjects in the same way; although one or more subject fields are recommended or optional in most institutions' metadata, more than 50% of records in the Digital Public Library of America (DPLA)[3] have no subject instances (Phillips & Tarver, 2021). However, one useful aspect of subject-based fields is that they do not rely solely on creation information or context (which is not always available for historic materials) and can be added or adjusted to improve a record as needed. This also renders subject-based research especially useful, as the collected data can directly inform the implementation of changes or distribution of resources.

---

[1] Optical Character Recognition

[2] This statement is based on research in progress; the analysis has been done, but not yet published. Raw data is available at ark:/67531/metadc1637688.

[3] https://dp.la/

1.1 Background

The University of North Texas (UNT) Libraries manage the Digital Collections,[4] which comprise more than 3.2 million items accessible from three public interfaces: The Portal to Texas History,[5] the UNT Digital Library,[6] and the Gateway to Oklahoma History.[7] Materials come from partner institutions across the states of Texas (for the Portal) and Oklahoma (for the Gateway), and from departments within UNT (for the Digital Library). These collections span a range of item types including images (such as photographs, drawings, postcards, and maps) and both printed and handwritten text (such as letters, technical reports, newspapers, pamphlets, and journal articles), as well as audio, video, datasets, and other objects.

All items are archivally stored in a single digital library system with a uniform UNTL metadata schema[8] that provides the same twenty-one field options and requires the same eight fields for all records.[9] This consistency has allowed for the development of tools and guidelines across the entire system, and frees up resources for quality assessment. UNTL records require two subjects per item record, and the `subject` field includes various controlled terms as well as free-text keywords, with qualifiers to appropriately label the value types.

Most collections in the Portal and the Digital Library are digitized at UNT and described by students who are trained to do metadata in various Libraries departments. However, staff or volunteers at partner institutions may also edit records in their collections. Although partners often provide information about the content and history of the items (to add descriptive details), this information may not be known in many cases. For example, the location, creation date, and identities of persons pictured in photographs are often not available, particularly for historical images or photos collected over time.

The University Photography Collection, which is accessible via both The Portal to Texas History[10] and the UNT Digital Library,[11] is managed under the auspices of the UNT Libraries Special Collections department. The entire physical collection is being digitized as an ongoing project with additions each year, but the online collection currently contains more than 30,000 photos related to the UNT community over time (1890-present). Images include buildings and campus views, school events and performances, portraits of students and faculty, and other photographs documenting daily life and historic moments at UNT. According to usage statistics,[12] items in this collection are viewed roughly 15,000 to 23,000 times per month. This means that the University Photography Collection contains diverse image content for the purposes of research and subject assignment, and is also a heavily-used collection that would benefit from strong descriptive metadata and clear guidance as the collection grows.

---

[4] https://digital2.library.unt.edu/search/
[5] https://texashistory.unt.edu/
[6] https://digital.library.unt.edu/
[7] https://gateway.okhistory.org/
[8] https://library.unt.edu/digital-projects-unit/metadata/input-guidelines-descriptive/
[9] https://library.unt.edu/digital-projects-unit/metadata/minimally-viable-records/
[10] https://texashistory.unt.edu/explore/collections/UNTPC/
[11] https://digital.library.unt.edu/explore/collections/UNTPC/
[12] https://texashistory.unt.edu/explore/collections/UNTPC/stats/

1.2 Research Project

Given the importance of metadata in general, and descriptions for images in particular, we wanted to determine how we could more effectively support useful subject assignment for cultural heritage collections in the Digital Collections. Most internal guidance regarding metadata creation and subject assignment tries to align with other broad, international guidelines, to leverage system functionality (without relying solely on existing infrastructure), and to reflect expected user needs. However, while some tangential qualitative input has been solicited regarding subject values -- from public users related to the development of UNT digital library interfaces (Hartman & Murray, 2010) and from metadata editors (Fox et al., 2019) -- we have not previously gathered data to inform metadata guidelines about *actual* user subject needs.

Aside from generating this new data, our research furthers other previous work within the UNT digital library system, including formatting of guidelines for metadata creators (Tarver, 2010), evaluation of subject values in DPLA and The Portal to Texas HIstory (Tarver et al., 2015; Phillips & Tarver, 2021), and analysis of networking statistics to assess subject values in the UNT Electronic Theses and Dissertations (Phillips, Tarver & Zavalina, 2019) and targeted collections (Phillips, Zavalina & Tarver, 2019). These previous publications focused on data analysis of existing information captured within the UNT system or in aggregated, harvestable metadata (i.e., DPLA), which this research intends to supplement with varied data sources.

This study was funded by a UNT Libraries' Dean's Innovation Grant to explore user opinions about keyword relevance for images in the Digital Collections.[13] We sought to answer the following research questions:

- What subject terms do users prefer to search (or assign) to find particular images?
- What subject terms do users feel are "appropriate" or "inappropriate" for specific images?
- What expectations do users have about search results based on specific subject terms?

To delineate a clear scope, the project focused on materials in the University Photography Collection. That collection is useful in this context because it contains only photographs, and the content spans a variety of topics (e.g., people, events, buildings, sports, etc.) from both historic and modern time periods.

---

[13] IRB-20-455, approved December 23, 2020

2. Methods

This study employed both qualitative and quantitative methods through three separate components to determine the kinds of keywords that users prefer for finding digital photographic images. Two of the methods involved research with participants in a virtual setting, starting with a publicly-released survey and followed by a user research activity with respondents who volunteered for additional participation. The researchers also evaluated existing data derived from the user logs in the UNT Digital Library and The Portal to Texas History.

Due to the COVID-19 pandemic at the time that we developed our research goals, we planned for all activities to occur online. Although similar research (e.g., user research activities) generally makes use of interactions with participants in person, this was the best way to ensure the safety of both researchers and participants, and also allowed for the possibility of a broader geographic range of respondents.

2.1 Survey

The researchers developed a survey[14] in Qualtrics related to images from the University Photography Collection that asked 12 questions in four main categories:

> (1) a single image with a free-text box for participants to provide keywords that they would use to search for the image
>
> (2) a single image with a list of keywords for participants to choose any that they might use to find similar or related images
>
> (3) a single image with several keywords for participants to rank according to relevance
>
> (4) a keyword with a series of images asking participants to choose any that they would expect to find when searching that keyword

Additionally, two multiple-choice questions asked how participants expected items to be labeled with organizational names that change over time. And, to allow grouping of final answers, five demographic questions asked participants to describe previous digital library usage, age range, and education completed. All questions were required in the interest of compiling complete, comparative data; however, the age and education questions included a "prefer not to answer" option.[15]

Participants were recruited online via e-mail, the UNT Libraries' newsletter, and social media posts, starting January 21, 2021. These methods employed some level of snowball sampling by asking readers to share the original posts or information with others who might want to take the survey. Links to the

---

[14] https://digital.library.unt.edu/ark:/67531/metadc1832874/

[15] Note that in some places, this paper references specific survey questions using the notation "Q#" where "#" refers to the number assigned in the survey text, and also matches the labels in the raw data.

survey included some basic embedded data so that we could track (roughly) where participants encountered the survey link.

We closed the survey on August 6, 2021 and had received 95 total responses.[16] Of these, 1 participant checked the box to acknowledge consent but did not answer any additional questions, leaving 6 partial (see Table 1) and 88 fully-completed survey responses for which we present data and findings.

Table 1. Number of questions (out of 19) answered in partial responses

Respondent	# Questions	Respondent	# Questions	Respondent	# Questions
A	10	C	1	E	3
B	10	D	1	F	1

2.2 User Research

User activities were conducted to gain additional information about user expectations for image keywords. While the survey was open, we collected contact information for 39 respondents who volunteered to participate in further research. Initial and follow-up emails were sent to the remaining respondents with details about the activities and the request of their preferred date and time for a 30-minute session. We also used social media to recruit additional participants. Volunteers that resided outside of the United States were not selected to participate.

These activities were completed using online software called Miro[17] while interacting over a Zoom[18] conference call. The purpose of the research activities was to:

- Determine the keywords users are most likely to use when conducting searches for photos in the University Photography Collection.
- Gain a better understanding of how photos should be described by staff from a user perspective.
- Gather information about how well the current keywords for collection photos match user expectations.

During the first activity, participants were asked to describe eight photos[19] from the University Photography Collection with 3-4 keywords of their choice. For the second activity, they were asked to describe another set of collection photos using a list of current keywords with the option to add additional terms. The participants were encouraged to think aloud during the process. Once both activities were complete, they were asked follow-up questions to provide them with the opportunity to

---

[16] Dataset: ark:/67531/metadc1833414
[17] https://miro.com/
[18] https://unt.zoom.us/
[19] Photos used in both activities and keywords provided in the second activity are in Appendix B. Some parts of this paper reference images with the notation "I#" where "#" matches the designation in Appendix B.

explain their keyword selections further and share any difficulties they experienced describing the photos.

2.3 Log Data Evaluation

The study also evaluated query terms used to find items in the University Photography Collection. Each time someone accesses an item in the public interfaces of the Digital Collections, the system logs the usage to track statistics and notes the referring URL (e.g., a search engine or internal site search). Additionally, when searches occur within the system -- in this case, the UNT Digital Library or The Portal to Texas History -- search terms are automatically appended to the end of the URL, connecting the queries and the items within the system logs. Queries may be initiated by a user typing in a system search box, or by clicking a linked value within a record (e.g., names, subjects, locations, etc.) to find additional items with the same creator, contributor, or publisher, or about the same topic or location, etc. No search terms are documented if the query originates outside the Digital Collections (e.g., from Google).

For the purposes of this study, a system administrator downloaded and processed the logs for all uses of items in the University Photography Collection (isolated by unique ARK[20] identifiers[21]) that occurred during January 1, 2019-December 31, 2020.[22] This resulted in a list of individual image identifiers and any known query terms that directed a user to each item (after eliminating any uses that did not include a query string). Data included 6,903 searches across 3,471 user sessions[23] and more than 3,000 unique search terms.

---

[20] Archival Resource Key
[21] https://texashistory.unt.edu/explore/collections/UNTPC/identifiers/
[22] At the time of this paper, the log data still requires some processing to make it shareable; the final dataset will be made publicly available in the UNT Data Repository (https://digital.library.unt.edu/explore/collections/UNTDRD/).
[23] User "sessions" were isolated based on IP address and timestamps; i.e., if a user from a single address viewed items within a brief timeframe or consecutive timestamps. These included cases where [1] a user searched multiple terms to find/view items and [2] a user searched a single term and then viewed multiple items from the results.

3. Respondents

One goal was to distribute the survey publicly to gain responses more representative of "general users" rather than library professionals or a specific user group. We included a handful of digital library usage and demographic questions at the end of the survey to get a sense of the participants who provided data.

First we asked if participants had previously viewed items in The Portal to Texas History or the UNT Digital Library, and then if they had previously viewed items in *any* digital library -- such as the Digital Public Library of America, Smithsonian Digital Library, New York Public Library, etc. -- (see Figure 1). Overall, most participants have viewed images in digital libraries before, including 50 who answered "yes" to both questions and only 5 who answered "no" to both questions.
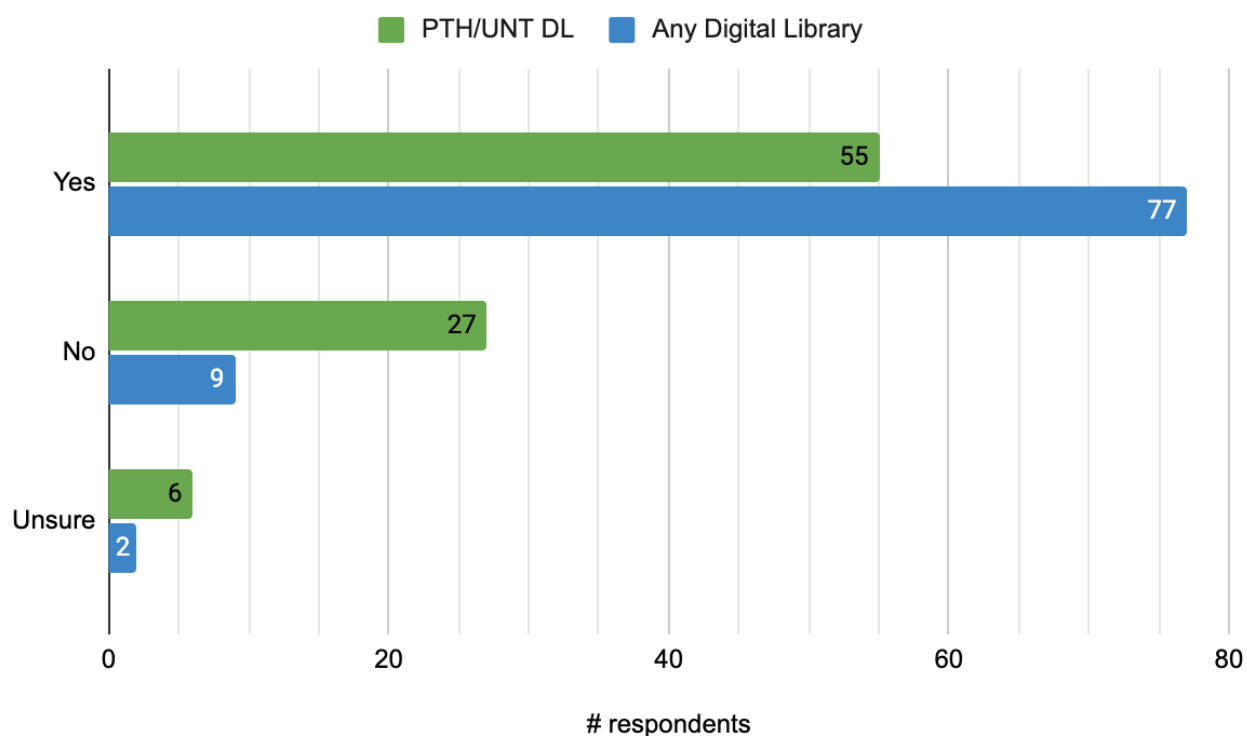

Figure 1. Number of respondents who have/have not viewed items in UNT systems or other digital libraries

As a follow-up, we asked *how* participants had used digital libraries in the past, with several choices and a write-in option for "other," which included 12 respondents who identified themselves as librarians of various specialties. Although this is not a good representation of the general population, it may be more representative of actual digital library users, which seems likely, given the number of respondents who said that they had used various digital library resources previously.

Based on the responses, a large number of survey participants regularly conduct research, or are in professions that would lend themselves to using digital archives to find information:

What was your primary role in searching/using digital libraries or archives?	
Academic researcher	20
General interest	16
Historian	10
Student	8
Genealogist	7
Other	20
Librarian	12

We also included two (optional) demographic questions for respondents to designate their ages (within decade ranges) and highest level of education completed; all participants who completed the survey provided an answer for these questions (see Figure 2).
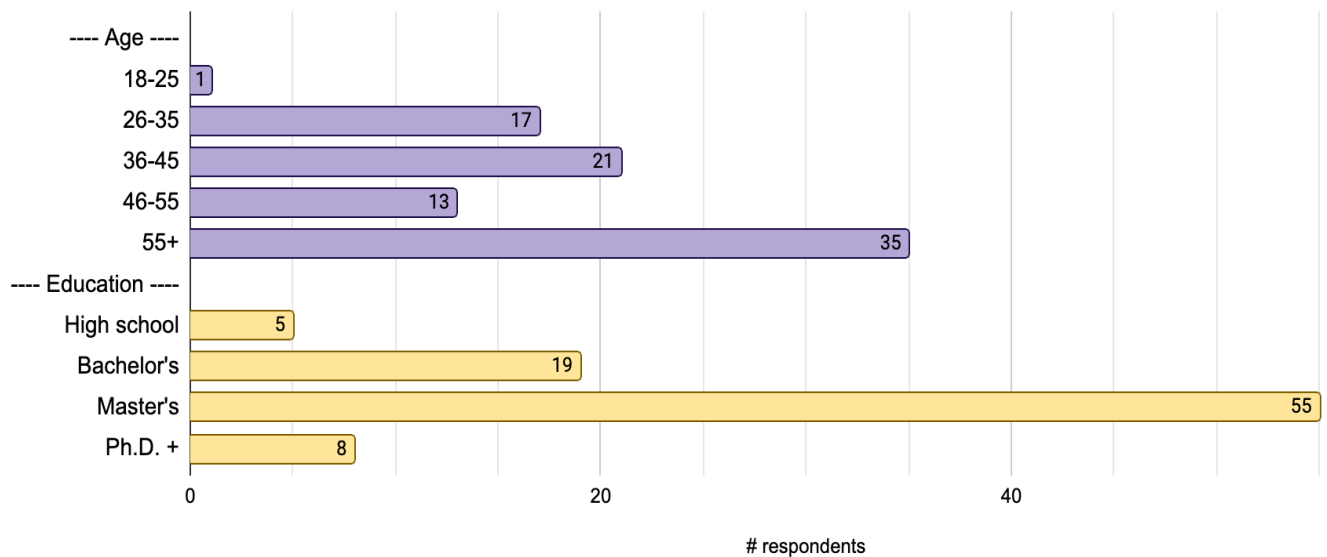
Figure 2. *Responses to demographic questions about age range and highest level of education*

Since we required participants to be at least 18 years of age, the ranges started with "18-25," though only one respondent identified as being part of that group. A majority of respondents identified themselves as being 56+ years old, with remaining respondents roughly split among the other three decade ranges. Additionally, a large majority of respondents -- 63 out of 87 (more than 72%) -- said that they had a Master's degree or above (Ph.D., post-doc, etc.).

4. Data Overview

Users have a variety of personal and professional research needs, as well as differing perspectives on terminology for searching. Although it is not possible to find perfect or "universal" terms to describe images, we were able to organize and categorize information collected during this study. Based on the various sources, there are a number of general trends related to the kinds of terms that users prefer or find most relevant for certain kinds of items. The next sections provide an initial analysis of the data.

4.1 Distribution of Information

For two images (see Figure 3), we asked respondents to suggest free-text search terms, to see what kinds of terms users might generate without guidance.

Figure 3. *Images used in questions soliciting free-text responses (Q3, left and Q8, right)[24]*

To keep questions simple, we asked respondents about possible search terms, but did not distinguish among different types of information that might be in other fields. For example, in the public interface for the Digital Collections, search results can be filtered based on values in a number of different fields (Krahmer, 2016). These filters overlap with some of the aspects that users suggested, such as the type of item (photographs); the city where the photo was taken -- i.e., the general location of the image -- (filterable by city, county, and state); and the date that the image was captured (filterable by decade, year, month, and day depending on the level of known specificity). Although all of the text in the records is indexed and findable as search terms, we would generally not repeat this information in the `subject` field since it is already included in more functional/filterable fields (see Table 2).

---

[24] Online images available at ark:/67531/metadc797285 (left) and ark:/67531/metadc949696 (right)

However, more specific or qualification terms might be appropriate as subject values. Although "photograph" would not be added (since it duplicates the `resource type` field), we would suggest including controlled terms or keywords that modify the type, e.g., "Aerial views" and/or "Aerial photographs" (from Library of Congress Genre/Form Terms) or "birds-eye images" (keyword). Similarly, names of places that are more specific than the city level (e.g., a neighborhood or park) -- in this case, the university, where these photos were taken -- would also be represented as searchable keywords or subject terms to supplement the filterable location. This distinction helps to isolate which suggestions may be useful for informing subject terms versus information that is better represented elsewhere.

Table 2. *Example values suggested by participants that would be parsed across different UNTL fields*

Category	Example Values	UNTL Field	Filterable?
Physical medium	**Photograph** [Location name]--**photographs** black and white **photography** of campus	`Resource type`	YES
	Black and white photograph **"black and white"** **black and white** photography of campus **Black and white**	`Description` (physical)	No
	aerial view **birds eye view** University of North Texas **panorama** **Aerial picture** **Panoramic view** of (name of city?)	`Subject`	No
Geographic	Photographs--[**city or neighborhood name**] **name of town** North Texas University in **Denton TX** aerial view of **(name of city?)** **the name of the place** if I knew it **Name of place.**	`Coverage` (place name)	YES
	Photographs--[**city or neighborhood name**] **North Texas University** in Denton TX **field house** graduation location **UNT Campus** graduation venues **(locations)** **UNT** **name of college/university**	`Subject` `Coverage` (place point or box)	No
Temporal	Graduation class of (**insert year**) **19xx's** architecture **1960s** **1950s 1960s** graduation ceremony **20th century**	`Date` (creation) `Coverage` (date) `Coverage` (time period)	YES

Among the suggested terms was the phrase "black and white" (or variations including b&w and grayscale), which was listed by 11 respondents for the aerial photo of the campus and by 6 respondents for the image of the student walking during graduation. Generally, we represent the color of photos as a characteristic of the item medium in the physical description, which includes the type of item, physical details, and dimensions, e.g., 1 photograph : b&w ; 8 x 10 in. While this text is searchable, it is not a filterable option and it would not be found searching "black and white." The number of times that participants included this phrase overall suggests that it is something that they would prefer to use, although this phrase only appeared once in all of the user activity responses. However, it is unclear why users listed this as a possible search term. For example, if it is meant to be representative of the age (i.e., "old" black-and-white photos versus "newer" color photos), then it may not be as useful as decades and historic time periods that are filterable. Additionally, images display in the search results and can be sorted by dates as well, so users can see which photos are color vs. black-and-white, but they would still have to know to add "b&w" to the search terms to *exclude* color images, if this is desired.

4.2 Number of Terms

Metadata editors often have questions about the number of subject terms that are appropriate to include in a record. To some degree, this will depend on the content of the image. For example, a studio portrait of an individual or a photo of a building facade provide limited opportunities to provide subjects beyond a name (if known), and general terms (e.g., "portraits" or "buildings"). While the context of the questions in the survey is not the same as completing actual metadata records, it is interesting to see the number of terms that individual respondents tended to choose or assign for particular questions.

In the free-text questions (accompanying the images in Figure 3), respondents could provide any number of search terms; some listed a single phrase or string of text, while others listed a series of words and phrases. For the purposes of this analysis, any words separated physically by line breaks or commas (in the response) were considered multiple terms, but otherwise text was considered a single search phrase. Generally, most respondents provided 1-3 terms (see Figure 4) although several responses included as many as 8 separated words or phrases.
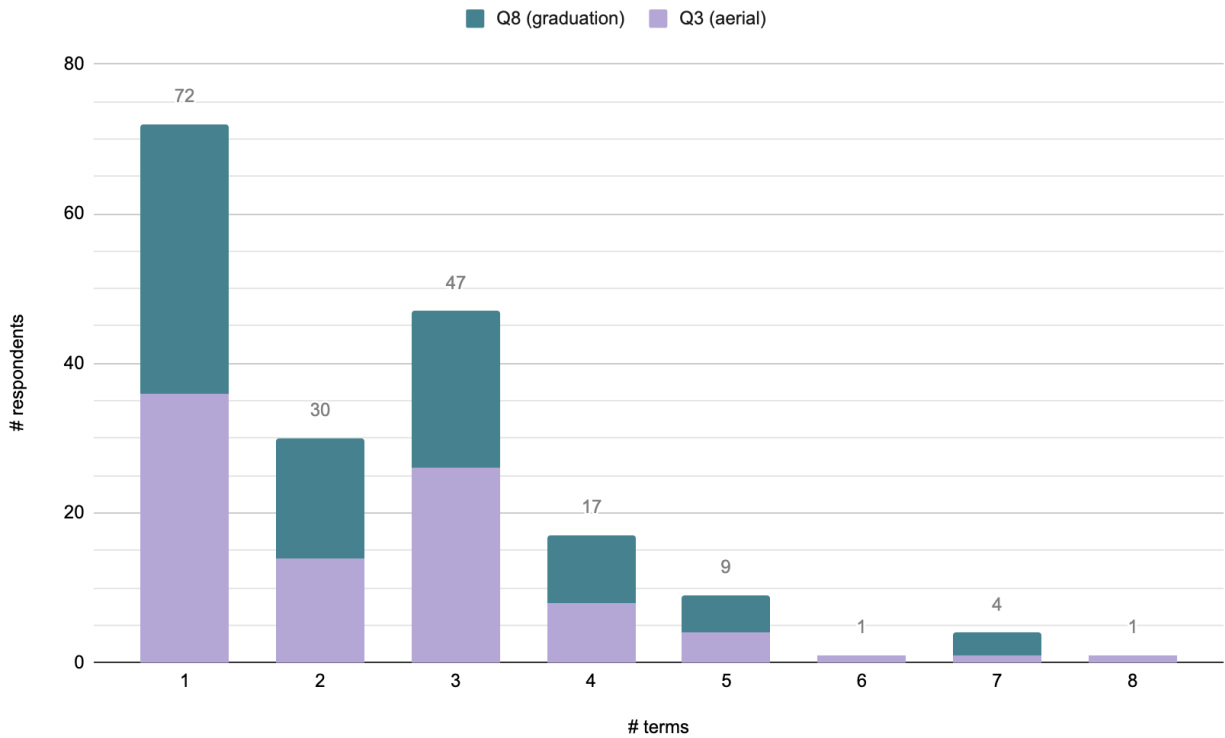
Figure 4. *Frequency of responses by number of parseable terms in free-text questions*

For three of the questions, respondents had the option to choose from a list of possible terms: questions 5 & 13 included 16 possible options (these used the aerial and graduation images from questions 3 & 8 respectively) while question 9 included 14 possible options to match with a group portrait of football players (see Figure 5).

Figure 5. *Images used for questions 5, 9, and 13 (left to right)*[25]

[25] Online images available at: ark:/67531/metadc797285 (Q5), ark:/67531/metadc1166507 (Q9), and ark:/67531/metadc949696 (Q13)

Participants tended to choose around 3-5 of the terms (see Figure 6). However, several respondents chose only 1 or 2, and one user chose every possible term for questions 9 and 13. The frequency with which respondents chose a particular number of terms tended to be roughly the same for each of the questions.
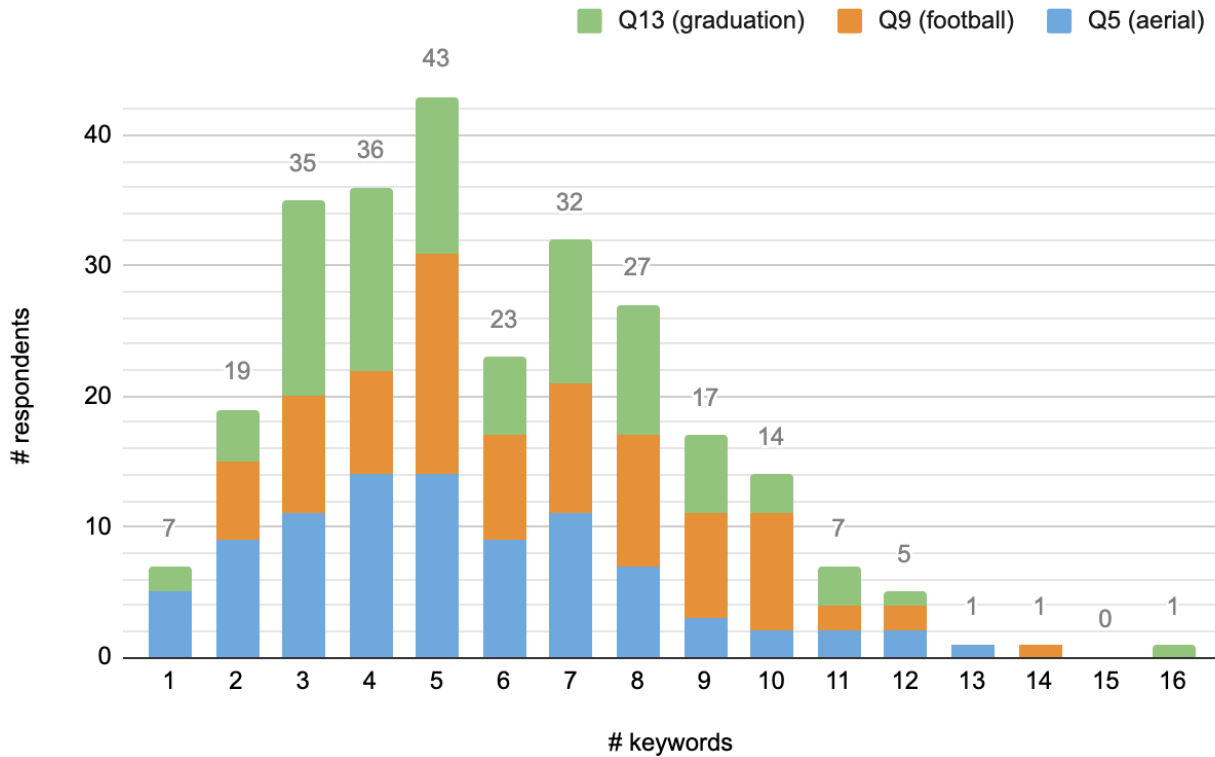

Figure 6. Frequency of responses by number of terms chosen per image

Of course, these numbers do not provide significant guidance for metadata creation. Records need to overlap with terms that many different users might choose including variations in spelling or terms (e.g., automobiles *and* cars). However, it is interesting that most participants only tended to choose 3-5 terms as the "most" relevant, even when offered a large number of possible terms.

5. Subject Analysis

Aside from terms that would be parsed into other fields,[26] there are several general classifications of topics that were chosen or suggested by respondents for finding information. Some of these include modifiers or terms that would add specificity to information in other fields. For example, to ensure consistency and browse options, place names are documented as hierarchies of administrative divisions -- e.g., United States - [state] - [county] County - [city] -- however there are scenarios where someone might want to search for units smaller than a city/town/village entity or "places" that do not fit in the hierarchy (e.g., New England). Similarly, specific *types* of photographs may be more easily represented as subject terms while still allowing users to find "all photographs" as a material type.

Responses to questions 5, 9, and 13 (chosen terms from a list) are summarized in Table 3, broken into five broad categories related to these modification-type subjects as well as item-specific topics:

1. **Type Modifiers** -- Any terms that describe the method or content of a type of item (in this context, photographs) e.g., tintypes, panoramic photos, portraits, etc. This category may not always be relevant.
2. **Places** -- More specific locations or areas that may not fit into an administrative hierarchy, e.g., a body of water, region, neighborhood, park, building, university campus, etc.
3. **People** -- Descriptors of types of people based on occupation, affiliation, gender, etc.
4. **General topics** -- Broad, browseable terms that would tend to return large numbers of items rather than particular images, e.g., science, architecture, business, agriculture, etc.
5. **Specific elements** -- Terms representing details in an image, dependent on the content (e.g., "fields" is a small detail in the aerial photo, but the *location* where the football image was taken)

These same categories of terms can also be used as a framework to organize the free-text subject terms in questions 3 & 8 (see Table 4) as well as the responses from the user research (see Tables 5 and 6). For the purposes of grouping similar terms, any term or phrase was counted if it included the particular keyword or variations. The number of common terms suggested by the free-text questions in the survey was still extremely low. Only two phrase terms -- aerial and graduation -- were suggested by more than 80 individual respondents. But the next-most-frequent (variations of the University of North Texas or North Texas State University) were only listed by 21 respondents for the aerial photograph and 3 respondents for the graduation photo. This means that fewer than 24% of users agreed on any other particular keyword phrase terms.

In the user research activities, more terms seemed to be commonly suggested, although the total number of participants was much lower (only 9 respondents). The second activity provided possible keywords, accounting for some of the similarity, though even in the first activity, participants tended to suggest similar terms. For example, roughly half of the subject terms for the first user activity (Table 5) were added to at least one image by five or more of the participants, though the overlap does increase for the second activity (Table 6).

---

[26] As outlined in Table 2.

Table 3. *Categories of keyword values chosen by participants in questions 5, 9, & 13*[27]

	Q5 (aerial)		Q9 (football)		Q13 (graduation)	
	Keywords	**#**	**Keywords**	**#**	**Keywords**	**#**
Type modifiers	aerial photographs	89	group portraits	53	photography	12
	photography	26				
Locations	university campuses	73	fields	11	gymnasiums gymnasia	27 14
	colleges	53	stadiums	3	stages	21
	schools	22			college campuses	16
	west side	2				
People			football players	89	graduates	74
			athletes	72	students	42
			students	33	audiences	32
					women	25
					people	10
					faculty	8
General topics	buildings	48	college football	71	commencement	91
	architecture	34	sports	55	graduation ceremonies	87
			athletics	48		
Specific elements	quads	31	uniforms	46	regalia	43
	administration buildings	27	footballs	45	walking	16
	roof tops	24	jerseys	27	basketball hoops	1
	streets	21	poses	18		
	parking lots	12	cleats	5		
	fields	7				
	cars	4				
	trees	2				

[27] Highlighted cells are values chosen by at least 60% (roughly 53) of the respondents

Table 4. *Categories of keyword values written in free-text boxes by participants in questions 3 & 8[28]*

	Q3 (aerial)		Q8 (graduation)	
	Term contains...	**#**	**Term contains...**	**#**
Type modifiers	aerial	89	black and white	6
	photograph/photography	12	photograph/photography	3
	black and white	9		
	birds eye/bird's eye/bird's-eye	5		
Locations	campus	41	college/university	12
	University of North Texas/UNT/ North Texas State University	21	University of North Texas/UNT/ North Texas State University/NTSU	3
	city/cities	18		
	town/downtown	10		
	college/university	10		
People			graduates	18
			student	15
			female/woman/women/girl	13
			crowd/audience/spectator	7
General topics	building	14	graduation	85
	historic	3	ceremony/ceremonies	32
			commencement	19
Specific elements	church/churches	4	stage/platform	9
	street	4	auditorium	7
	quad	3	gymnasium/gym	6
	urban	3	cap and gown/cap/gown	6
			walk	4

---

[28] Highlighted cells are values chosen by at least 60% (roughly 53) of the respondents

Table 5. Categories of keyword values added by participants in the free-text section of the user research[29]

	Term contains...	# responses for each image							
		1	2	3	4	5	6	7	8
Type modifiers	black and white					1			
Locations	campus					8		6	
	University of North Texas/UNT/ North Texas State University	4 2	1	2	4	5		5	3
	college/university	1			1	3		2	1
People	student		1	1		7		8	3
	alumni musician				8 --				-- 2
	female/woman/women/girl/lady							1	1
	pilot cheerleader			2 --	-- 6				
General topics	athletic/sport	1			2				1
	military aviation/flight		7 --	-- 3					
	event		4		2	2	2	1	1
	historical/vintage	1	2	3			1		
	architecture/building					4		1	
	social		6			1	2	1	
Specific elements	football	8							1
	WW2/WWII		4				3		
	uniform/fashion		3				2		
	airplane/plane movie/theater/theater			8 --			-- 6		
	memorabilia dance	2 --	-- 7						
	band/drum bike/bicycle							-- 7	9 --
	clock tower/clocktower/tower/bell tower					4		3	

[29] Highlighted cells are values chosen by at least 60% (roughly 5) of the respondents. Empty cells = 0 responses.

Table 6. *Categories of keyword values chosen or added by participants in the second section of the user research*[30]

	Term contains...	# responses for each image							
		9	10	11	12	13	14	15	16
Type modifiers	[none]								
Locations	campus	4			3			4	
	University of North Texas/UNT	5		1	2	1	7	7	3
	town/downtown/square field/stadium		5 --						-- 6
	college/university				1				1
People	students	8		5	7	1		6	
	female/women/women's/girl			3		3			
	faculty/staff/teacher	2			1				
	cheerleader		4						1
	athlete/player			7		5	(1)[31]		2
	crowd/audience/spectator/attendee				4		1	5	2
General topics	athletic/sport			1		1			1
	event	4	5				2	1	
	interior/exterior	1	1		1	1	1	1	
Specific elements	parade/float		9				3	3	
	football basketball			-- 9					8 --
	uniform courthouse		-- 4				8 --		
	classroom/lecture hall	1			5				
	dribble/dribbling band/marching band/instrument			8 --			-- 9		
	lecture	1			9				
	flag							1	8

---

[30] Highlighted cells are values chosen by at least 60% (roughly 5) of the respondents. Empty cells = 0 responses.
[31] One respondent added the term "Players" for this image, referencing musicians, rather than athletes.

5.1 Specificity of Terms

The most common terms chosen by participants in the multiple-option questions[32] tended to be broad: commencement, aerial photographs, football players, graduation ceremonies, graduates, university campuses, athletes, college football. Even among free-text responses, where there was less total frequency, the most common terms (or phrase components) were: aerial, graduation, campus, ceremony/ceremonies. Similarly, in questions matching images to a search term, respondents tended to choose terms that represented major components, rather than "anything visible" (see Tables 7 and 8).[33]

Table 7. *Responses regarding images that participants expect to find in question 2*

Search term: "shirts"			
Image	**Responses**	**Image**	**Responses**
	10		49
	91		0
	16		94

---

[32] As outlined in Table 3.

[33] Highlighted cells in Tables 7 & 8 are values chosen by at least 60% (roughly 53) of the respondents. See Appendix A for links to the online images.

Table 8. *Responses regarding images that participants expect to find in questions 6 and 11*

Search term: "architecture"		Search term: "college campuses"	
Image	**Responses**	**Image**	**Responses**
	84		0
	90		86
	0		6
	87		87
	0		29

This tendency for participants to choose general terms is most noticeable in the images that did not receive *any* responses to the question, "Which of these images would you expect to see in the search results for the keyword ["keyword"]?" Each of the questions had at least one image that was not chosen by any respondent; in each case it was an image for which the term was *technically* accurate, but not *obviously* accurate or *as applicable* as the match with the other images. This suggests that even when users tend to err on the side of considering a wide range of specific terms as "relevant" for a particular image -- or when terms describe less-central elements -- these users may not actually want or intend to use those terms to *find* images. Comparatively, in the first user research activity, participants tended to write in specific terms[34] related to the images (e.g., UNT clock tower) rather than broad terms (e.g., architecture). Even in the second user research activity, where participants had a list of existing terms to assign, there was a marked tendency toward more specific terms.[35]
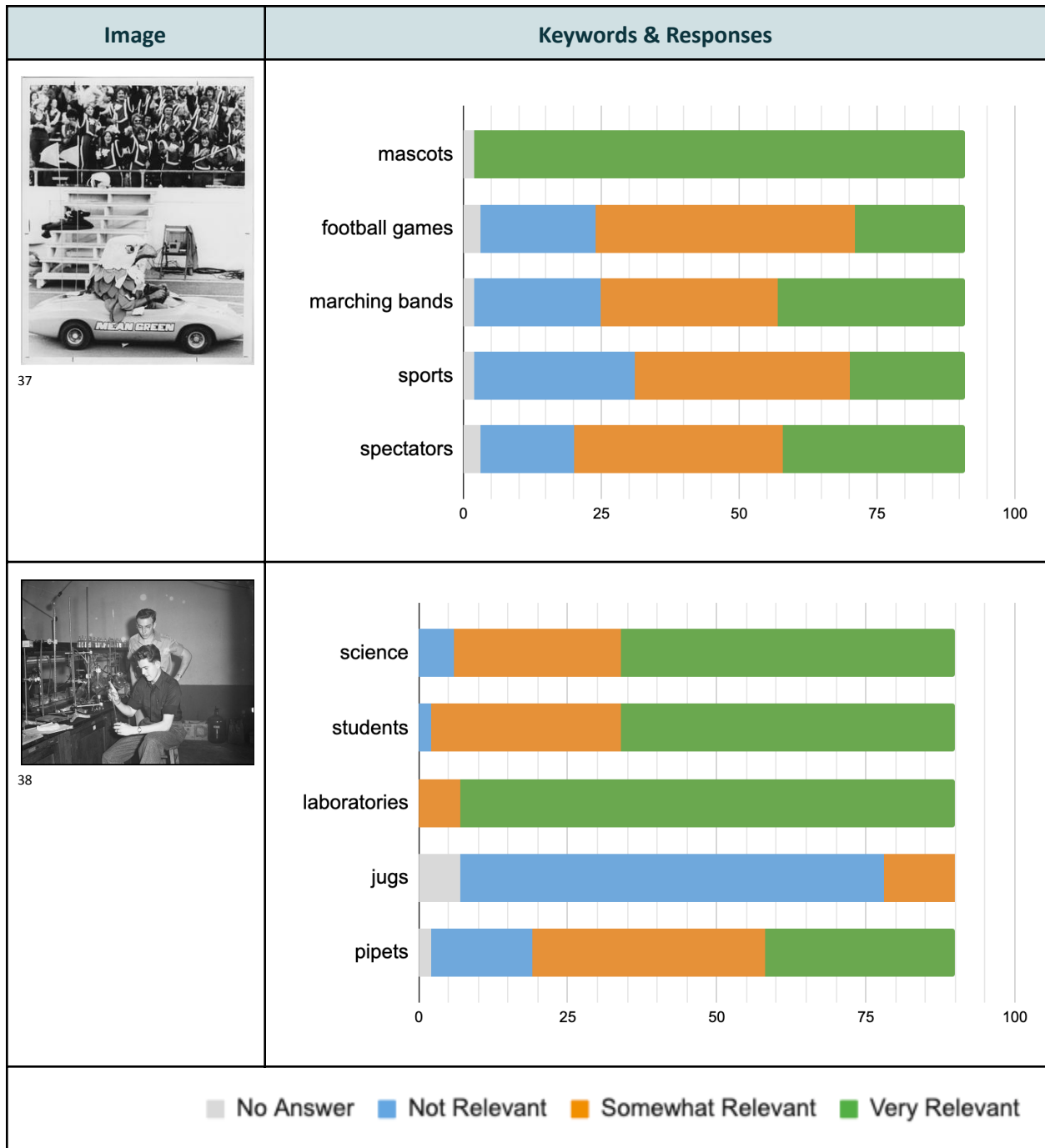
Survey questions 4, 7, 10, and 12 provided a list of five possible keyword terms along with an image and asked respondents to rank each keyword on a scale of 1 (not relevant) to 3 (very relevant).[36] Responses for these questions displayed a much wider range of acceptable "relevance" than the multiple selection questions. Aside from a couple of outliers, many of the terms were nearly-evenly split among not relevant, somewhat relevant, and very relevant responses (see Tables 9 and 10). Responses that rated each term *at least* somewhat relevant (i.e., total responses ranking either 2 *or* 3) include roughly half of the possible terms for each image. Possible terms included both the general (science) and extremely specific (cellos), although the level of perceived relevance varied depending on the image; for example, all respondents marked the specific term "mascots" as *very relevant* (Q4), but 79% of respondents considered "jugs" to be *not relevant* (Q7).

---

[34] As outlined in Table 5, with summary data in Table 11.
[35] As outlined in Table 6.
[36] Note that although question responses were required, in some cases, users did not mark a ranking for a specific keyword in the list. By default, this was represented in the data as a ranking of "0."

Table 9. *Frequency of relevance rankings for terms, with images (questions 4 and 7)*

Image	Keywords & Responses
 37	
 38	

Legend: No Answer (gray), Not Relevant (blue), Somewhat Relevant (orange), Very Relevant (green)

Table 10. *Frequency of relevance rankings for terms, with images (questions 10 and 12)*

Image	Keywords & Responses
  39	
  40	

No Answer · Not Relevant · Somewhat Relevant · Very Relevant
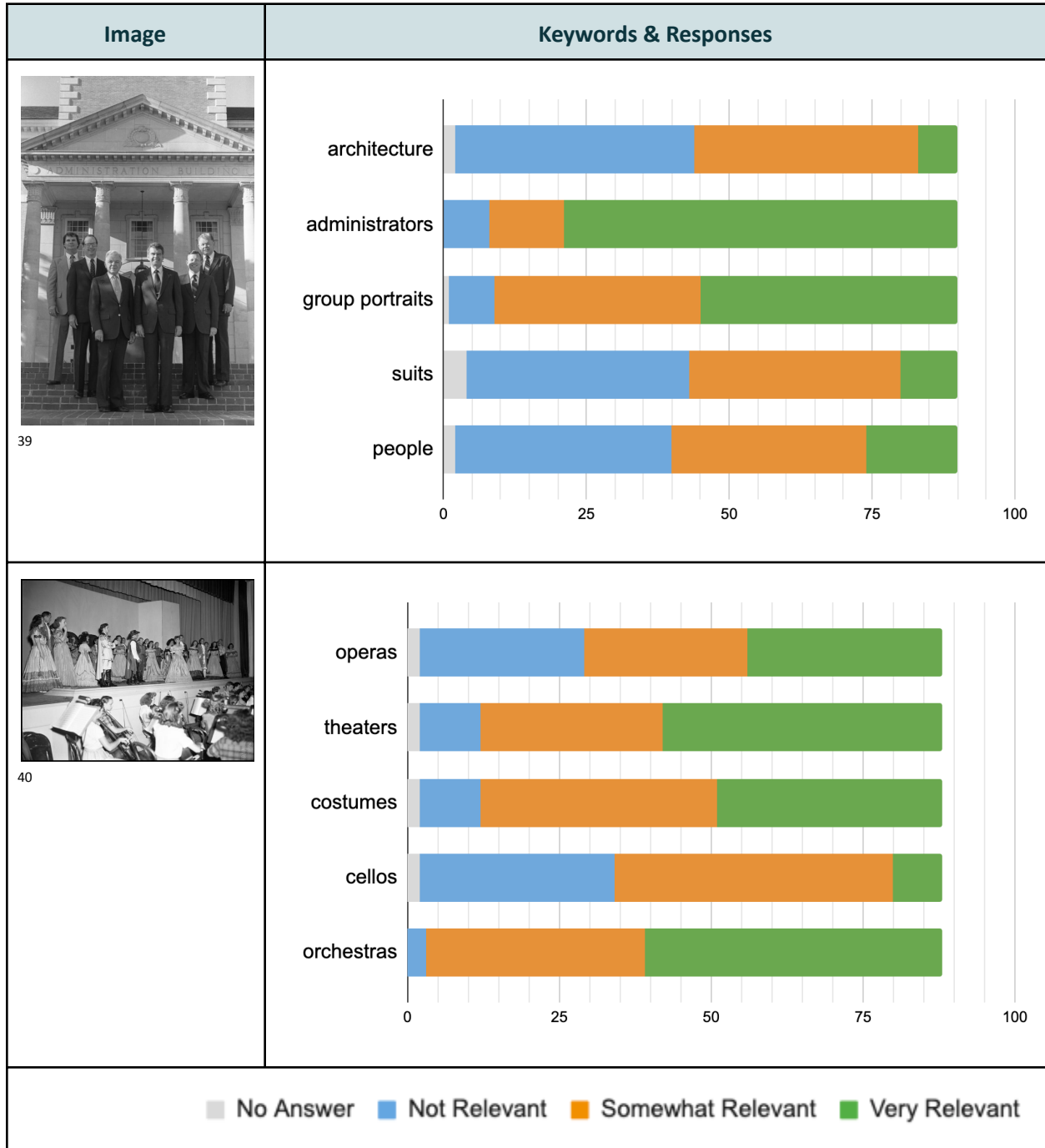
39 Image online at: ark:/67531/metadc1151842
40 Image online at: ark:/67531/metadc179448

5.2 Organization Names

Name changes are common among colleges and universities, as well as other corporate bodies, but can pose issues in the description of historical photographs. For example, the University of North Texas was established by Joshua Chilton in 1890 as Texas Normal College and Teachers Training Institute and has had seven different formal names from its founding to present day. When a user searches for historical images related to this university, they may struggle to determine if they should search by the current name or the name the university had at the time the materials were created. This issue may also arise in records regarding individuals who may have legally changed their name or gone by various names over their lifetime.

1890-1894   Texas Normal College and Teachers Training Institute

1894-1901   North Texas Normal College

1901-1923   North Texas State Normal College

1923-1949   North Texas State Teachers College

1949-1961   North Texas State College

1961-1988   North Texas State University

1988-   University of North Texas

Standards for archival description indicate that when titling an element, it is appropriate to "use the last (latest) name of the corporate body in the materials being described" or, optionally, to "use the name under which the bulk of the material was created" (Society of Southwest Archivists, 2021). Internally, UNT Libraries also has suggestions for how current and historical names should be represented in various metadata fields, although it is not always consistent in practice. Within the survey, we inquired about *users'* expectations in order to determine if our approach in this specific area is relevant or helpful to users when navigating digital library records by the university's name.

Two questions in the survey asked whether participants would expect to find *all items* related to the entire history of an organization, or *only items* from the time period that matches a particular version of the name based on [1] a historic name or [2] the current name (see Figure 7). Overall, most respondents (53%)[41] tended to expect that searching for the **current** organizational name should find everything, but that a **historic** name should only find items created at the corresponding time period. However, aside from the general consensus, not all respondents agreed, including some respondents who wanted to find *all items* regardless of which version is used, and a handful who wanted the opposite: expecting that items should always match the time period of the name (even the current version).

---

[41] 37 plus 3 "other" responses that primarily correspond with the same categories (e.g., find everything *and* also have references or sorting by time period is still counted as "*all items over time*") for a total of 40 out of 76 respondents who answered this question.
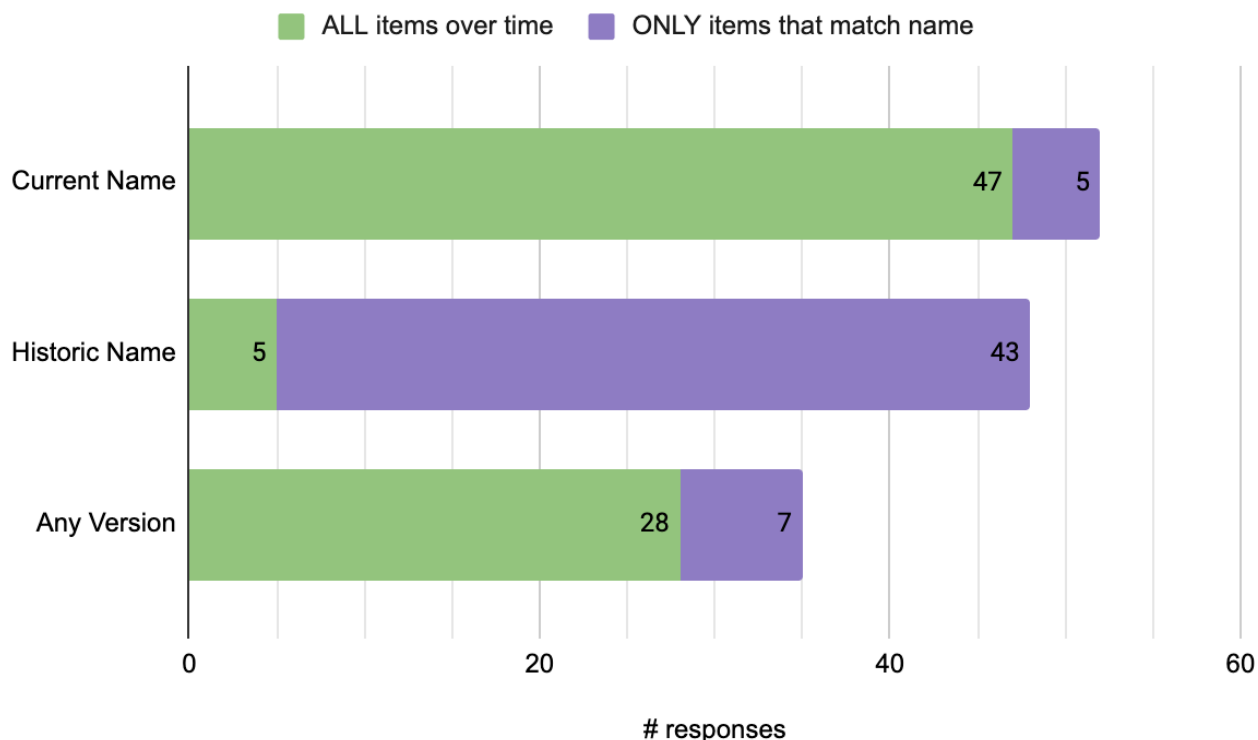
Figure 7. *Responses regarding how participants expect historic organization names to affect search results*

Facilitating these different approaches is a difficult balance -- and does not even fully account for images in the collection for which we have no specific date (and may not be able to determine which historic name would apply). As previously noted, for items that do have dates, both dates and broad time periods[42] can be used to filter search results, even if organizational names are not always perfectly consistent, or do not exactly match differing user preferences. This also aligns with some user preferences, including a write-in response that explicitly clarified an expectation that search results "have [a] feature of sorting by time period" in order to find images from particular date ranges, regardless of the organizational name at the time an image was taken or the terms a user searches.

---

[42] Controlled list of Texas history eras: https://digital2.library.unt.edu/vocabularies/coverage-eras/. These time periods (as well as decades, years, and custom date ranges) can be browsed in the public interface across the Portal (https://texashistory.unt.edu/explore/dates/#time-periods), selected within the University Photography Collection (https://texashistory.unt.edu/explore/collections/UNTPC/dates/), or used to filter results after a keyword search.

6. Users vs. Metadata Creators

While not explicitly stated in the research questions, comparing the free-text terms supplied by participants (in the survey and the first user research activity) to the current subjects assigned to the same images may help to highlight any major differences or gaps between expectations and practice. Subjects support item findability in two broad ways: (1) as part of the "full text" of the indexed record, and (2) as an exact match for "all of the items containing [term]." The second scenario would generally apply if a user chose to "explore by subjects"[43] and selected a subject browse term, or if a user viewed a metadata record and clicked on a specific term to launch a search for that precise subject value. For the purposes of this analysis, we are assuming that a user would simply type the free-text terms into a search box (either a search engine, or the search box in the Digital Collections), which would search each word as an individual query term.[44] Hence, a partial match against any of the existing subjects would retrieve the image, even if the exact free-text phrase does not appear in the record.

The free-text questions include survey questions 3 and 8 (see Table 12) and the first user research activity (see Table 13). For each of the images, free-text and existing subjects are highlighted to denote partial and full matches -- i.e., search terms that could retrieve the image that the users were describing. Additionally, Table 11 provides a summary of how many free-text terms matched, broken out into the general subject categories. Overall, terms representing specific elements tended to match most often, followed by locations, and then people. General topics were only matches for roughly half of the free-text terms; type modifiers almost never matched (however, aside from the aerial/bird's-eye photo, there were extremely few "type" terms suggested, as these subjects are not always relevant).

Table 11. *Summary of # partial or full search term matches |vs.| # unique participant free-text answers[45]*

	Q3	Q8	I1	I2	I3	I4	I5	I6	I7	I8
Type modifiers	2 | 7	1 | 3	--	--	--	--	0 | 1	--	--	--
Locations	5 | 10	4 | 6	4 | 5	--	--	2 | 2	4 | 5	--	4 | 5	2 | 2
People	--	1 | 9	--	1 | 1	1 | 3	7 | 8	4 | 4	0 | 4	6 | 7	2 | 5
General topics	1 | 2	1 | 4	3 | 7	3 | 9	2 | 5	2 | 4	2 | 7	0 | 8	1 | 2	1 | 2
Specific elements	1 | 5	1 | 9	5 | 9	10 | 12	3 | 14	4 | 6	9 | 12	0 | 12	5 | 8	11 | 13
TOTAL[46]	9 | 24	8 | 31	15 | 21	14 | 22	6 | 22	15 | 20	20 | 29	0 | 24	17 | 22	17 | 22

---

[43] This is an option in the Portal (https://texashistory.unt.edu/explore/subjects/) since those records are required to have at least one term from the local hierarchical vocabulary (https://digital2.library.unt.edu/subjects/browse/). Subjects cannot be browsed in the Digital Library, which does not apply a uniform subject vocabulary.

[44] Technically, a general, search-box query would match against text anywhere in the record (including the title and content description), but we are focusing solely on subjects to keep the analysis straightforward.

[45] Highlighted cells represent a match of 60% or more between free-text values and existing subjects.

[46] Total query returns: Q3--38%, Q8--26%, I1--71%, I2--64%, I3--27%, I4--75%, I5--69%, I6--0%, I7--77%, I8--77%

For the survey images (Table 12), one mis-match in question 3 was that the record includes specific street and building names, which participants did not have; alternately, some participants misunderstood the context, adding "cities" or "downtown" or even "churches" after mistaking the Hurley Administration Building clock tower for a steeple. In question 5, all of the existing terms were matched, but a number of suggested terms (including "graduation" and "ceremonies") are not in the record.

Table 12. *Comparison of free-text subject terms in vs. existing terms in the online record in survey questions 3 & 8*

	Participant-Suggested Terms	Subjects in Online Record
Q3	aerial photograph/photography black and white birds eye/bird's eye/bird's-eye campus University of North Texas/UNT/ North Texas State University college/university city/cities town/downtown building historic street quad church/churches urban	University of North Texas -- Photographs. Education - Colleges and Universities - University of North Texas Landscape and Nature - Aerials Architecture - Buildings Architecture - Civil Works - Streets and Roads Highland Street Maple Street Avenue C Quadrangle Dormitories dormitories dorms residence halls Crumley Hall Music Annex Men's Building Laboratory School Journalism Building Women's Gymnasium Lab School Gymnasium Kendall Hall Education-Home Economics Building NTSU North Texas State University UNT college campuses
Q8	photograph/photography black and white college/university University of North Texas/UNT/ North Texas State University/NTSU graduates student female/woman/women/girl crowd/audience/spectator commencement graduation ceremony/ceremonies gymnasium/gym stage/platform auditorium cap and gown/cap/gown walk	University of North Texas -- Photographs. Education - Colleges and Universities - University of North Texas UNT college campuses commencements graduates gymnasiums

Match results for images in the first user research activity (Table 13) were more mixed than the survey questions, although there were fewer total participants and the existing records have varying levels of subject assignment to search against. In fact, the record for image 6 contains only six subject terms and, although the number is not a direct correlation to matches (e.g., Q8), there were *no* matches to participant terms, so the few existing terms are not aligning with expectations. However, ¾ of the images have a retrieval rate of 60% or higher, compared to 25-38% for the two survey questions (see Table 11). Also, for this activity, participants tended to suggest fewer terms related to locations and people, and there was only a single instance of a type modifier term among all eight images (I5).

Table 13. Comparison of free-text subject terms in vs. existing terms in the online record in the user research

	Participant-Suggested Terms	Subjects in Online Record
I1	North Texas State/North Texas State University UNT University/College Athletics Football/UNT football Footballs Promotional items Memorabilia Vintage Team Special collection Historical football Football memorabilia Football prizes First State Bank/1st State Bank Green and white football UNT toy football	Education - Colleges and Universities - University of North Texas Sports and Recreation - Football exhibits footballs athletics branded items souvenirs McCain, Fred NTSU First State Bank
I2	UNT students Military Interiors Fashions WW2/WWII Dance Historical Social Groups Social gathering Military ball Social event WWII Social Event Military social Military dance Dance event Couple dance Uniforms Vintage military uniforms UNT ROTC 1940s dance	Education - Colleges and Universities - University of North Texas People Social Life and Customs Military and War - Personnel Military and War - Wars - World War II Education - Schools - Students exhibits guests dances dancing soldiers College campuses North Texas State Teachers College

Table 13 (cont.). Comparison of free-text subject terms in vs. existing terms in the online record in the user research

	Participant-Suggested Terms	Subjects in Online Record
I3	UNT students Pilot/Pilots Aviation Aircraft Airplane/Plane Historical Vintage airplane Historical airplane 2 seater airplane Take off Test Flight instruction Hand crank Flight/Flights/flying Prop Turning airplane propeller Parachute Propeller	University of North Texas -- Photographs. Education - Colleges and Universities - University of North Texas Education - Schools - Students Business, Economics and Finance - Transportation - Aviation - Airplanes Military and War - Wars - World War II UNT college campuses North Texas State Teachers College NTSTC
I4	University/College Cheerleader/Cheerleaders Alumni UNT cheerleader UNT Cheer Alumni UNT alumni Cheer alumni Acrobat Sports/Athletics Athletic events Cheerleading UNT cheering North Texas Cheer Alumni Balance Alumni event [name of stunt] Cheer performance	Education - Colleges and Universities - University of North Texas Business, Economics and Finance - Service Industries - Photography Education - Schools - Cheerleaders Sports and Recreation - Cheerleading Education - Alumni People - Individuals Education - Events - Homecoming Education - Schools - Students Fouts Field performances lights t-shirts UNT stunts

Table 13 (cont.). *Comparison of free-text subject terms in vs. existing terms in the online record in the user research*

	Participant-Suggested Terms	Subjects in Online Record
I5	Black and white College/University University of North Texas UNT campus Campus UNT students People relaxing Student/Students Campus life Campus buildings Students gathering/Student gathering Groups Gathering Exteriors University outdoor event Campus architecture UNT social event UNT students on campus Campus Speaker Clocktower/Bell Tower UNT clock tower/UNT clocktower UNT bell tower UNT Admin Building Administrative building	Education - Colleges and Universities - University of North Texas Business, Economics and Finance - Service Industries - Photography Education - Events People - Individuals Education - Schools - Students Hurley Administration Building Administration Building UNT University Day Library Mall stages crowds groups bands exteriors trees clocktowers performers
I6	US Military members/personnel Navy man with lady Soldier WW2/WWII Military Historical Fashion Movie theaters/Movie theatre Theater Military life Social events Movie going Ticket booth Military formal Movie WWII era social event 1940s military uniforms Consession stand Military couple Uniforms US Guard	University of North Texas -- Photographs. Education - Colleges and Universities - University of North Texas UNT college campuses World War II World War II Recorded History Collection

Table 13 (cont.). Comparison of free-text subject terms in vs. existing terms in the online record in the user research

	Participant-Suggested Terms	Subjects in Online Record
I7	University of North Texas College/University UNT campus/Campus Students/UNT students UNT students on campus College students Lady with bicycle Student with bike People relaxing Student life Campus buildings Student gathering UNT quad Bicycle Backpack UNT social event Biking on campus UNT clocktower/Bell Tower	Education - Colleges and Universities - University of North Texas Business, Economics and Finance - Service Industries - Photography People - Groups Arts and Crafts - Music - Bands People - Individuals Education - Schools - Students Education - Events University Day Library Mall stages crowds Hurley Administration Building Administration Building UNT exteriors trees clocktowers performers bikes backpacks Bicycles.
I8	College/University Students/UNT students Female musician Musicians Tympany drum player Band Athletic events Game UNT football Percussionists Band percussion Student band Marching band Xylophone Drumline Drum/Drums UNT Marching band--percussion section Marching band percussion Outside band practice	Education - Colleges and Universities - University of North Texas Business, Economics and Finance - Service Industries - Photography Arts and Crafts - Music - Marching Bands People - Individuals Education - Schools - Students UNT moves Fouts Field fields members Mean Green Brigade Mean Green Marching Band practices band members xylophones pits mallets percussion t-shirts drums drumline trumpets horns

Overall, many of the existing terms that did not match user suggestions were specific names (events, buildings, streets, etc.), which are important to users searching for UNT history, but understandably not information that general public participants would have. In fact, when asked to provide feedback, almost all of the participants in the user research activities reflected that they did not have context, or

needed more context, or were unclear about specific aspects of the images that (presumably) they would otherwise have described. Although context is often known for *this* collection, metadata editors often face the same problem if specific information is not available. Even when contextual information *is* known, subject assignment needs to reflect generic image content for users who do not know names when searching.

6.1 Logged Searches

For comparison, we reviewed data from actual user searches. These included a large number of extremely specific terms, such as dates and names of particular persons, buildings, cities, events, streets, etc. compared to participant suggestions. Additionally, a number of search terms reflected precise photographs, including titles and unique archival identifiers, rather than topics. The most commonly-searched terms still include a number of broad topics (see Table 14[47]), although they may be paired with more limiting terms such as dates. These differences are likely due in part to knowledge of Denton and UNT, but also the fact that actual users may have had images in mind or use cases to match to images, versus the research participants who did not have any context.

Actual searches also included quoted terms (i.e., searches to find "more like" a particular image by clicking on a value in a record or selecting from the public browse interface[48]) -- some of which were dates, time periods, and possibly creators (name queries are not labeled as subjects or creators; some of the searches could be either). Quoted terms also represented general subject values and 301 different searches for 63 unique hierarchical subject terms from the UNT Libraries Brows Subjects (UNTL-BS) that allow browsing in the Portal.[49]

Aside from the previously-discussed issues (e.g., representing all possible synonyms of topics), another common concert in metadata creation can be spelling variations. This is problematic for historical collections where names may not be spelled consistently in source documents, but even "known" names could be difficult for users if they are unaware of the "correct" spelling or preferred version. Similarly, multiple searches in the log data included the misspelling "arial" when looking for "aerial photographs." Although it may not be possible to cover every possible avenue, each data point in this case did represent a user looking at a photograph, so the searches did return *something* worth reviewing, at least.

---

[47] Data was organized by item usage, but this table reflects "how often" a term was searched, e.g., multiple item uses found with the same search term in a single session were counted as a single "search"
[48] https://texashistory.unt.edu/explore/ (Portal) or https://digital.library.unt.edu/explore/ (Digital Library)
[49] https://texashistory.unt.edu/explore/subjects/

Table 14. Categories of commonly-searched keyword values in queries by public users, 2019-2020[50]

	Search Term	#	Example query strings
Type modifiers	photo/photograph	132	1920 panoramic photograph; yearbook photography; truck photo
	aerial	95	1919 aerial; denton college aerial; apogee stadium aerial
Locations	north texas	279	drinking fountain north texas; wooten hall north texas
	university college	270 81	world war ii university of north texas; university college inn; denton normal college; college of industrial arts
	ntsu unt	52 265	ntsu golf; ntsu spirit; ntsu logo denton; ntsu student studying unt transportation; unt police; unt 1890; unt cheerleaders
	denton	298	denton square; normal school denton; nike missile denton
	campus	77	unt campus 1921; campus chat; campus map
	normal school	59 58	north texas normal school pennant; normal building denton laboratory school; school boys; elementary school
People	women woman	63 32	unt women's gymnasium; women vote; women basketball tall woman; woman shopping
	student	121	ntsu student; student studying; students on campus; student life
General topics	football	187	unt football student section; depot football; football player 84
	basketball	61	basketball player; unt varsity basketball; women's basketball bloomers
	building	277	unt main building; unidentified building; historical building
	music band	167 136	music building; musicians; music practice hall marching band; two o'clock lab band female
Specific elements	-- hall	180	bruce hall; crumley hall; stovall hall; kerr hall unt
	war	95	1940 texas war; consumer rationing in world war two; war bonds
	willis	83	a. m. willis; brad willis; unt willis; willis library exterior
	library/libraries	134	willis library; human library; unt library mall
	scrappy	47	unt scrappy; scrappy mascot; scrappy the eagle
	homecoming valentine	69 47	unt homecoming; homecoming queen; homecoming scrappy valentine's day; valentine
	union	113	student union; unt union piano; university union
	department	47	unt psychology department; department store; department of english

[50] Based on preliminary data; counts may not reflect final analysis. Search values were normalized for analysis.

7. Conclusions

The results of this initial analysis are mixed, but we can draw some broad conclusions. Depending on the context of the question -- i.e., "is this applicable/relevant" vs. "find items *like* this one" -- both general and specific subjects may be considered appropriate. This also lines up with the idea that specific terms may be most useful to find *particular* images but general subjects could be used to browse entire categories of images or materials, meaning that a combination of subjects types is most helpful for users. This is also supported by the comparison of free-text suggestions and existing subjects. Users often included more general terms and type modifier terms (e.g., "birds-eye views") than metadata editors. Those terms represented the categories least likely to return search results and suggest a need for combinations of terms at different levels of specificity.

In general, the terms that respondents wrote in (questions 3 & 8) reinforced the preference for users to search or filter by names, places, and dates/time periods that has previously been established during user testing of the Portal (Murray & Belden, 2010). Unfortunately, for historic materials -- particularly photographs -- some or all of this information is often unknown. The fact that users *want* to include specifics and are also frustrated by a lack of context is reflected both by write-in survey answers (e.g., "the name of the place if I knew it") and in comments made during user activities -- such as, "Need more context about photos" and "There is no context, so it is difficult to describe the photos. The only choice is to be vague." Within the scope of the University Photography Collection, most locations and time periods are known (or may be reasonably inferred), but some of these details may not be represented consistently across the collection, which may also impact user experiences.

Respondents seemed to prefer (or more often agree upon) broad, general topics, which are important for browsing; however, this may also reflect some of the difficulty of providing terms specific enough to find particular images when names, dates, and places are not known -- by users or by curators -- which occurs even in many real-world scenarios. The lack of agreement in free-text responses also illustrates the challenge for metadata creators to include enough different terms that will meet the search expectations for a large number of users. For example, one participant in the user activities commented that "Some photo descriptions are too specific, and some are too vague. There is a lack of consistency."

Based on the expected results of participant-suggested terms (Tables 11-13), there do appear to be some significant gaps in the way that subjects are currently assigned to photos in this collection. For example, the fact that there were *no* matches for I6,[51] even among broad topics, and that even when some terms matched, records may be missing terms of significance, such as "graduation" for Q8[52] (although "commencement" is in the record and matched some user suggestions).

---

[51] As noted in Table 13.
[52] As noted in Table 12.

7.1 Next Steps

One purpose of this research is to inform future metadata creation to make images more findable for users. Based on these initial results, we can make several recommendations for editors that may be incorporated into the existing subject guidelines.[53]

> **Include subjects at multiple levels.** Users need both specific terms and general topical terms to account for the amount of knowledge available to a searcher (e.g., a user may want to find images of campus buildings or events, but not know the name of a *particular* building/event) or preferences (e.g., whether they are browsing for "any" image within certain parameters or looking for an image of a specific person, building, event, etc.).
>
> **Use a variety of subject types.** A useful framework *may* be to include at least one subject from each of the relevant "categories" identified in §5 -- i.e., type modifiers, locations, people, general topics, and specific elements.
>
> **Focus on primary content.** Rather than naming everything that is visible in the image, users seem most concerned about terms that are relevant to the main content (i.e., what the image is "about"), while also including any unique elements that might help to find particular images within a large collection. This could complement or help direct the assignment of different subject "types."

Ultimately, although we can work to improve subject assignment, there will always be limitations. It would be unreasonable for metadata editors to include all possible variations of subjects, and we tend to discourage duplication of information across fields in most cases -- which is sufficient for general keyword searches, though not specific subject searches. Users also have different needs and ways of interacting with collections (e.g., browsing for general topics versus searching for specific topics or names), which require multiple kinds of subjects. Finally, even taking all of these aspects into account, editors can never describe things in a way that will be ideal for every user, as demonstrated by the variety of preferences regarding historic name representation. Despite these challenges, this data and the recommendations will help to inform ongoing metadata creation and hopefully improve overall findability.

---

[53] https://library.unt.edu/digital-projects-unit/metadata/fields/subject/

References

Fox, N. T., Tarver, H., Phillips, M. E. (2019). *Identifying Gaps in Tools and Interfaces for Assessing Metadata Quality* [White Paper]. https://digital.library.unt.edu/ark:/67531/metadc1453742/

Gentry, S., Hale, M. L., Payant, A., Tarver, H., White, R., & Wittmann, R. (2020, May). *Survey of benchmarks in metadata quality: Initial findings* [White Paper]. Digital Library Federation, Assessment Interest Group, Metadata Working Group, Benchmarks Sub-Group. https://digital.library.unt.edu/ark:/67531/metadc1637685/

Hartman, C. N. & Murray, K. R. (2010, February). *Optimizing the User Experience in a Rapid Development Framework: Final Project Report*. https://digital.library.unt.edu/ark:/67531/metadc32974/

Krahmer, A. (2016, May). *User Experience of Access Points: Eye-tracking, Metadata, and Usability Testing* [Doctoral dissertation, Texas Tech University]. https://digital.library.unt.edu/ark:/67531/metadc919593/

Murray, K. R. & Belden, D. (2010, May). Applying user-centered design principles to redesign the interface to the Portal to Texas History: The IOGENE Project. *Proceedings of the Society for Imaging Science and Technology Archiving Conference.* https://digital.library.unt.edu/ark:/67531/metadc28324/

Phillips, M. E., Tarver, H., Zavalina, O. L. (2019). Using Metadata Record Graphs to Understand Controlled Vocabulary and Keyword Usage for Subject Representation in the UNT Theses and Dissertations Collection. *Proceedings of the Electronic Theses and Dissertations Conference* (1st ed., vol. 2019, pp. 61-76). Associação Portuguesa de Biblioteconomia, Arquivística e Documentação. https://digital.library.unt.edu/ark:/67531/metadc1616586/

Phillips, M. E., Zavalina, O. L., & Tarver, H. (2019, September). Using Metadata Record Graphs to understand digital library metadata, *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 49-58. https://digital.library.unt.edu/ark:/67531/metadc1616551/

Phillips, M. E. & Tarver, H. (2021). Investigating the use of Metadata Record Graphs to analyze subject headings in the Digital Public Library of America, *The Electronic Library*. [in press]

Society of American Archivists' Technical Subcommittee. (2021). "Section 2.3 Title." *DACS (Describing Archives: A Content Standard)*. https://saa-ts-dacs.github.io/dacs/06_part_I/03_chapter_02/03_title.html

Tarver, H. (2010). Better Guidelines, Better Functionality: How Metadata Supports the Cycle of System Improvement at the University of North Texas. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 165-176. https://digital.library.unt.edu/ark:/67531/metadc29323/

Tarver, H., Phillips, M. E., Zavalina, O. L., Kizhakkethil, P. (2015). An Exploratory Analysis of Subject Metadata in the Digital Public Library of America. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 30-40. https://digital.library.unt.edu/ark:/67531/metadc725779/

United States Office of Education. (1930). *Survey of land-grant colleges and universities, Volume 1.* Government Printing Office. https://digital.library.unt.edu/ark:/67531/metadc276271/(m1/637/)

Appendix A: Images from Questions 2, 6, 11

This appendix includes links to the online images for the survey questions (2, 6, and 11) that asked: "Which of the following images would you expect to find by searching the term '[term]'? (Choose all that apply)."

Results for the responses to these survey questions are in §5.1 (Tables 7 and 8).

Images from Q2, search term: "shirts"		
ark:/67531/metadc850882	ark:/67531/metadc1635711	ark:/67531/metadc1633385
ark:/67531/metadc1635657	ark:/67531/metadc856284	ark:/67531/metadc1633361

Images from Q6, search term: "architecture"		
ark:/67531/metadc179540	ark:/67531/metadc1635666	ark:/67531/metadc177485
ark:/67531/metadc797719	ark:/67531/metadc1636525	

Images from Q11, search term: "college campuses"		
ark:/67531/metadc493625	ark:/67531/metadc797901	ark:/67531/metadc1639415
ark:/67531/metadc289229	ark:/67531/metadc1073531	

Appendix B: User Research Images

The user research activities included two parts; the first provided images and asked participants to add free-text keywords. The second set of images also allowed free-text keywords, but included a list of existing keywords (taken from the item records) that participants could match to images as well.

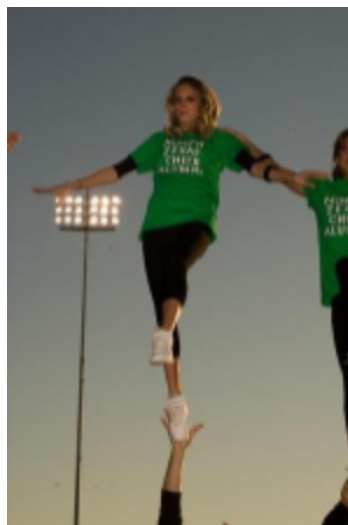| Part 1: Images soliciting free-text responses |
|---|

54

55

56

57

---

[54] Image 1: ark:/67531/metadc1633319
[55] Image 2: ark:/67531/metadc1633318
[56] Image 3: ark:/67531/metadc282370
[57] Image 4: ark:/67531/metadc1636580

58

59

60

61

---

58 Image 5: ark:/67531/metadc1635263
59 Image 6: ark:/67531/metadc282530
60 Image 7: ark:/67531/metadc1636591
61 Image 8: ark:/67531/metadc1636078

Part 2: Images with available keywords & free-text options					
Existing keywords					
UNT A&M flag teams meets pools	tracks floats crowds events police courts	eagles fields sedans swimmer viewers parades	windows players jackets athletes opponents teammates	dribbling UNT campus competitors flag runners photography cheerleaders	photographers student athletes Northern Arizona University Denton County Courthouse Eagle Student Services Centers Florida International University
Images 9-12					

---

[62] Image 9: ark:/67531/metadc979161
[63] Image 10: ark:/67531/metadc177473
[64] Image 11: ark:/67531/metadc1636533
[65] Image 12: ark:/67531/metadc1633422

Part 2 (cont.): Images 13-16

66

67

68

69

---

66 Image 13: ark:/67531/metadc1635414
67 Image 14: ark:/67531/metadc978398
68 Image 15: ark:/67531/metadc1636696
69 Image 16: ark:/67531/metadc1635718