

Utah State University

DigitalCommons@USU

Library Faculty & Staff Publications

Libraries

1-30-2024

The Path of Least Resistance: Optimizing Metadata Practices Through User Assessment

Andrea Payant
Utah State University

Liz Woolcott
Utah State University

Paul Daybell
Utah State University

Becky Skeen
Utah State University

Anna-Maria Arnljots
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/lib_pubs

 Part of the [Library and Information Science Commons](#)

Recommended Citation

Andrea Payant, Liz Woolcott, Paul Daybell, Becky Skeen, Anna-Maria Arnljots, (2024). The Path of Least Resistance: Optimizing Metadata Practices through User Assessment. *Journal of Library Metadata*. DOI: <https://doi.org/10.1080/19386389.2024.2303872>

This Article is brought to you for free and open access by the Libraries at DigitalCommons@USU. It has been accepted for inclusion in Library Faculty & Staff Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



Article Title:

The Path of Least Resistance: Optimizing Metadata Practices Through User Assessment

Author Details

Andrea Payant^{a*}, Liz Woolcott^a, Paul Daybell^a, Becky Skeen^a, & Anna-Maria Arnljots^a

^a*University Libraries, Utah State University, Logan, Utah, USA*

*Corresponding author: Andrea Payant, andrea.payant@usu.edu, Phone: 435-797-4343

Andrea Payant <https://orcid.org/0000-0001-9873-1538>

Liz Woolcott <https://orcid.org/0000-0002-6017-1392>

Paul Daybell <https://orcid.org/0000-0001-9830-5870>

Becky Skeen <https://orcid.org/0000-0002-0215-6936>

Abstract

As part of a multi-faceted research project examining user engagement with various types of descriptive metadata to improve metadata services, Utah State University Libraries Cataloging and Metadata Services unit investigated user search behavior in library catalog (MARC) records, Encoded Archival Description (EAD) finding aids, and most recently in digital collection metadata. The authors used web log analysis to determine how users interacted with CONTENTdm, categories of search terms used, and where search terms were found in a record. Key findings included that navigation through CONTENTdm using clickable queries (links) is a prominent user search pattern, dates are an important faceting tool and date ranges as a clickable link are heavily used, users demonstrated a more engaged research pattern when searching with geographic terms, and subject and transcription metadata fields match user search terms at a significant rate and are likely to be high drivers of search results.

Keywords

CONTENTdm, User-Search Behavior, Web Log Analysis, Dublin Core Metadata

Introduction

USU Libraries Cataloging and Metadata Services (CMS) unit has been engaged in a multi-year research project involving the study of user search behavior when looking for resources in library collections. The goal of the research has been to improve discoverability and workflows based on user needs gleaned from web log analysis. The first phase of the research focused on how users engaged with the online library catalog and MARC records exclusively (Woolcott et al., 2021). The second phase of the research centered on user engagement with varying levels of description in Encoded Archival Description (EAD) finding aids (Daybell et al., 2022). This paper concentrates on the third phase of this research, which is to examine user engagement with digital collections and study the discoverability of the Dublin Core metadata produced within CMS.

The aim of this work was to answer the following questions relating to user engagement within digital collections: How are users interacting with our digital asset management system, CONTENTdm? What search terms are used most? And where are search terms found in the metadata record? Finding answers to these questions gave the unit greater insight on user needs to optimistically improve the discoverability of online resources. Google Analytics data and web log analysis has helped give insight into the pros and cons of CONTENTdm, how users search our collections, user engagement with controlled vocabularies when searching, and what metadata fields aid most in user searches.

Literature Review

Biswas and Marchesoni (2016) analyzed usage data of CONTENTdm using Google Analytics to identify their institution's most visited collections. The gathered data was used primarily to demonstrate the value of harvesting through sites like the Digital Public Library of America (DPLA). They also show how usage information can inform decision-making for meeting user needs when building collections. The authors examined item views, referrals, sessions, and looked at keyword search terms to determine the subject matter for the most used collections. Their methodology within Google Analytics included determining item views as a percentage of total views, item referrals as a percentage of total referrals, and the number of items as a total of all items for all collections. They uploaded keyword search data from Google Analytics to OpenRefine then sorted, counted, and clustered it to determine the most used terms. They defined sessions based on geographic location, percentage of item referrals from search engines and DPLA, and new and returning visitors by collection. Their data revealed volume of use, access, and user location which were indicative of user interests. The data also showed that most users were engaged with collections that were not related to their institution but that repeat visitors engaged more with institution-specific collections, which implies the necessity of considering the needs of both the public and the institution. The authors identify the need for the use of qualitative data in addition to quantitative data to assess user needs. They also discuss using keyword search data from transaction logs to add to collection metadata as local subjects and further study if this impacts usage.

Therrell (2019) discusses descriptive levels in digital image collections. In opposition to the recent theory of More Product, Less Process (MPLP) posited by Greene and Meisner (2005), Therrell presents data that indicates that lower description levels in digital collections leads to negative impacts on discoverability. The author analyzed information seeking behaviors for images from the New York Public Library (NYPL) digital collection and on Flickr. The descriptive terms that were used to describe the items and the success rate in retrieving resources using those terms were examined. The methodology included vocabulary analysis and testing retrieval. Vocabulary analysis involved finding and listing descriptive terms, combining similar tags, and counting frequency. Retrieval testing entailed conducting searches and inverse searches for a selection of items within the collection. Findings for the vocabulary analysis showed how closely user-generated terms correlated with existing controlled vocabulary terms. For example, natural language tags in Flickr, while corresponding somewhat with Library of Congress Subject Headings (LCSH), were more variable and had broader scope than the more specific LCSH terms on the NYPL site. Findings for the retrieval analysis, particularly with the inverse searches of LCSH terms on Flickr and tags on the NYPL site, showed that using natural language tags in Flickr led to increased retrieval than using LCSH on the NYPL site. Results also showed that one item described only at the collection level was particularly difficult to retrieve, the lack of metadata being the reason for this difficulty. Therrell acknowledged that both collection and item level descriptions are useful in retrieval but argues that at least some description at the item level for digital image collections is necessary for successful retrieval and increased user satisfaction.

Gvianishvili et al. (2010) analyzed user behavior in a large, open-source digital library system used by scientific and academic communities. The sequential actions performed by users for more than 130,000

queries were captured and then statistical calculations were made later to determine the number of page views (abstracts viewed), downloads, displays (record listed on search results), and sees (record being seen). The user log data was used to rank and analyze the relationship between records. Results of the data analysis showed the most used search terms and their frequency. Data also indicated that sorting or adjustment to the display of material is rarely used, that users prefer to control search results through the search terms and typing, and that the advanced search is not used very frequently. The authors suggest that the resulting data can be used to accomplish their goal to focus on the specific needs of the user to improve access to information by applying a new ranking algorithm, building recommendation systems, identifying user communities with common interests, query expansion, interface optimization, and optimizing most requested queries.

Zavalina and Vassilieva (2014) report on the results of a study of user search logs in two large-scale, discipline specific digital libraries. Google Analytics data from these repositories were made publicly available and utilized by the authors to analyze and compare user search queries during a one-year period. Transaction logs showed patterns for user searching. The user queries were coded and categorized. Data analysis across both platforms showed that users preferred basic keyword searching to advanced searching but that the frequency of advanced searching in these cases were higher than with similar studies, indicating that users with discipline specific knowledge are more likely to use and advanced search. Results also revealed that users differed significantly between the systems when comparing query variation, length, overall frequency, and frequency of search category occurrences. With regards to usage based on the categories that were assigned, the authors reached the conclusion that geographical and personal names were prevalent with humanities users and object searches also occurred frequently. Authors suggest that faceted search options and search limitations will better serve STEM communities and that more granular item attributes like persons or places would most benefit cultural heritage collection users. They also emphasize the importance of advanced search options in digital library development as their findings show a clear user need, regardless of discipline.

Han and Wolfram (2016) examined usage patterns with digital image collections in CONTENTdm by gathering transaction logs from clickstream data. Their methodology included network analysis, sequential pattern mining, and *k*-means cluster analysis. They identified eleven types of user interactions by analyzing URL requests, comparing them to CONTENTdm and its navigational structure. Adjacent to findings of this study, they found that users engaged mostly in focused and direct browsing and simple search related activities with very limited viewing of individual items within collections. The author's analysis also indicated little to no user engagement with search actions, like facets. This underutilization suggests that these features aren't essential for user interactions or that there may be a need for further instruction on their use. They advise further research in other digital image collections to confirm precedence of their conclusions, specifically the uniformity shown in user search behavior.

Like the close look at user search terms in this article, White and Radio (2020) sought to determine if user search terms aligned with metadata present to describe datasets within a GeoBlacklight (geospatial) repository with a goal to aid metadata creation and optimize their system. The authors employed query log analysis, categorization of logs, and metadata field mapping as a methodology to reach their conclusions. Utilizing Google Analytics and Python scripts, they were able to acquire data and map search phrases. They found that among all query types, search phrases mapped mostly to the description and title fields within the metadata. They also concluded that topical searches mapped twice as often to the description field than to the subject field, suggesting a disconnect between user search terms and terms used in controlled vocabularies meant to describe the aboutness of a given dataset. Therefore, the authors emphasize the importance of thorough descriptions for improved discoverability. They also recommend more cohesive use of metadata standards among geospatial database practitioners and conducting similar research in other GeoBlacklight systems to compare results.

In a project similar to this study's methodology, research goals, and research questions, Hanrath and Radio (2017) discuss Google Analytics data gathered for users of an Institutional Repository. User search data for queries combined with successful searches resulting in downloaded content was collected. A random sampling of the results were further analyzed and compared to Faceted Application Subject Terminology (FAST) to determine if users were searching topically. The queries were also compared against metadata values for title and subject using string matching. Findings showed that there was a correlation between user search terms and FAST and that topical searches made up a significant portion of queries. Their data also showed that successful searches matched FAST at a higher rate than with metadata values.

In another study of user search behavior, Bogaard et al. (2019) used log analysis compared to metadata records in a historical newspaper collection to identify search patterns to help improve the user interface, search algorithms and collection management. Differing from this paper's analysis, the authors focus on facets and clicked results and not search queries. The main takeaways were that faceted searches were more common than non-faceted searches and search patterns can be identified correlating with specific aspects of a collection, like time periods, geographical locations, and subject.

Han, Joo, and Wolfram (2014) discuss using transaction logs and click-through data from an image-based collection in CONTENTdm to determine the types of search actions users engage in, how search actions are utilized, what queries and terms are used, and if there are differences between queries that are internal and external to the digital collection system. The authors found that internal queries and terms showed consistent relationships with each other, were more subject driven, and indicated use of controlled vocabularies. On the other hand, external queries did not reflect consistent relationships, were longer, and more diverse. Generally, queries were found to be short, only about 2 terms, and personal names and geographic locations were the most common searches. They also concluded that users spent more time evaluating search results and items than they did formulating or re-formulating their searches. The authors suggest that the evaluative behaviors prevalent in their findings should encourage system changes, like displaying more metadata information in search results lists, to reduce the need to click on individual items to see more information when browsing.

A review of literature shows that studies of user behavior analysis can lead to evidence-based insights into how metadata is utilized and ways to make meaningful changes to improve user interactions with digital collections. Literature also suggests that web log analysis is an effective way to gauge user behavior. These conclusions support the goal of this research project to gather web log data to analyze user search behavior within our institution's digital collections and find ways to improve user experience by optimizing metadata practices.

Background

USU Libraries' digital collections, known as Digital History Collections, are primarily comprised of digitized Special Collections & Archives (SCA) materials. SCA collections are organized under the following areas: Art and design, folklore, government information, manuscripts, photographs, rare books, and university archives. There are presently 57,087 items in 148 published Digital History Collections. The collections are approximately comprised of 48% photographs, 26% manuscripts, 14% Oral Histories, with the remaining content mainly consisting of books, poetry, art, and newspapers.

USU Libraries utilizes CONTENTdm as a digital asset management system. The metadata schema used within CONTENTdm is qualified Dublin Core. Metadata at the collection level only displays to users on landing pages created for each collection. When browsing or searching, users see limited item-level metadata next to each entry. When users click on specific items within a results list or visit the item

record page directly, they are presented with a full qualified Dublin Core record that also includes local fields. If the item happens to be a multi-file item (such as a book with multiple pages), then a compound item record is displayed first followed by an item-level record for each page. Both record structures are qualified Dublin Core with local fields, but instructional policy guides metadata creators to fill out the compound object record with the complete Dublin Core metadata and only include the page title and transcript (if available) for the item level record.

The Mountain West Digital Library (MWDL) harvests USU's Digital History Collection content and makes it publicly available on their discovery portal. MWDL is a service hub of the Digital Public Library of America (DPLA), which further aggregates our content on their site. As a result, we follow the metadata standards currently outlined in version three of the MWDL metadata application profile or MAP (Mountain West Digital Library Metadata Application Profile Task Force, 2019), which includes recent guidelines set by the DPLA. The locally published MAP for CONTENTdm collections at USU Libraries incorporates these regional and national metadata standards, which are applied whenever creating metadata for Digital History Collections (Payant et al., 2020). The MAP only requires the title field for minimum record viability, but in local practice, all records typically are built with a core set of fields containing the Dublin Core mapped fields of title, description, creator, subject (LCSH controlled vocabulary), publisher, date, coverage-temporal, language, source, rights, type, format and identifier, as well as the fields of subject (local keywords), physical collection information, call number, collection inventory (URL), uploaded by, metadata by, ARK ID, and ARK URL. File specifications such as checksum, color space, and pixel information are recorded automatically in the metadata upon upload, but are not supplied by the metadata practitioner. For subject fields, a minimum of one LCSH-based term is required. And where possible or applicable, contributor, medium, coverage-spatial and transcript fields are highly encouraged. In total, there are 39 possible fields, before allowing for repeatable fields. Typically records have a minimum of 21 fields that are populated by metadata practitioner, sans the file specifications.

Methodology

To gather the data needed to address the questions posited by the research team, Google Analytics was used to compile a dataset of web logs. Google Analytics gauges user behavior for CONTENTdm collections through tracking the number of pages viewed, the amount of time users spend on the site, and the percentage of single-page sessions. The research team selected three days from which to gather web logs. The three days were chosen purposefully to gauge user engagement during the mid-semester period, on a Tuesday, Saturday, and Thursday. The day selection included two weekdays and one weekend day to capture a generally representative sample across an entire week. The study was limited to three days to allow capacity for the research team to hand-code and analyze the web log results. The pageviews per day for each of the three days measured were 1,354 (Tuesday), 286 (Saturday), and 1,326 (Thursday). These averaged to 988 views per day. This is within range of the average daily pageviews in a year which is 933 per day.

A customized report was created for each day to include the links accessed within the set 24-hour period, the time spent on a page, and pages views. Logs of URLs accessed with the library staff IP addresses were removed from the results using a specialized view within the system to gather data more representative of digital collection users without including library staff actions. Over all three days there were 2,900 URLs accessed by users. These reports were exported and put into Airtable, which is a relational database employed by USU Libraries.

Octoparse, a web scraping tool, was then used to scrape user search results pages from CONTENTdm URLs gathered from Google Analytics. The Octoparse dashboard allows users to copy and paste lists of

URLs to be scraped for content. An auto-detect feature recognizes web page elements that can be previewed and selected and customized columns of data can be specified to produce data output. Once parameters are set the program will run a scrape for all the listed URLs. Using this capability, the user search results pages from CONTENTdm digital collections were scraped systematically for items listed on search results pages, facets present on search results pages, and for specific elements on each item record page. The web scraped data was then input back into Airtable, where the data was coded one by one for facets (order number and Dublin Core field), order of search results, records displayed in a search, and search terms input by the user. To improve inter-coding reliability, the coding process was performed by two individuals working separately and the resulting data was checked by a third person for anomalies.

Results and Analysis

Question 1: How are users interacting with CONTENTdm?

1.1 What pages are most frequently accessed in CONTENTdm?

There are six main types of pages coded from the logs of CONTENTdm URLs that users accessed over the course of the research project. They included:

Record Page: Webpage that includes the digital object and associated metadata.

Search Page: Webpage that lists the results of a search query, with results being the links and brief views of record pages that match the search query.

Landing Page: Opening webpage for a collection that includes information about a digital collection and may also include pre-built links to record or search pages regarding specific topics or grouping of digital objects in the collection.

Home Page: The primary or starting webpage for USU’s CONTENTdm collection. This page lists all the collections that can be found in the database.

Advanced Search: Search query page that allows for more granular searching, giving options for selecting specific collections, metadata fields, or date ranges.

Other: Pages that do not fit into the previously mentioned categories, including the “about” page and the “login” page.

Category of Webpage Accessed	Pageviews	Percent of Total
Record Page	2007	60.4%
Search Page	997	30.0%
Landing Page	270	8.1%
Home Page	41	1.2%
Other	5	0.2%
Advanced Search	3	0.1%
Sum	3323	100.0%

Table 1. Pageviews by webpage category

As shown in Table 1, the record pages were the most likely pages to be accessed in the database, representing just over 60% of all URL pageviews. The next most common category was the search page category, representing 30% of the pageviews. Landing pages for collections represented only 8.1% of pageviews, with the CONTENTdm homepage bringing in 1.2% of pageviews. The Other and Advanced Searches categories commanded only less than 1% of pageviews each.

1.2 How many record page views were the result of a search?

Many content management systems report the search terms or query conditions in the URL for a record page when the view of that record page is the results of a search query. However, CONTENTdm does not record the URL in the same fashion. Therefore, when a user clicks on an item in the results list, the resolving record will elicit the same URL as someone who clicks on a Google or browser search. Therefore, the research team could not specify the exact number of records that were accessed because of a search within the system. Instead, the research team examined how many records pages appeared or did not appear in search results lists to approximate the maximum number of times record pages *could have been* accessed from within the system versus from an external source.

Record Pages	Pageviews	Percent of Pageviews
Accessed AND in search results	236	11.8%
Accessed but NOT in search results	1771	88.2%
Sum	2007	100.0%

Table 2. Pageviews by accessed records that appeared in search results and accessed records that never appeared in search results.

As Table 2 demonstrates, record pages were significantly more likely to be accessed but never appear in search results, with 88.2% of all records accessed by a user never appearing in search results. This indicates that most users were viewing a record after clicking on a link from a browser search or using a direct link.

1.3 How often is the advanced search feature used?

CONTENTdm's advanced search offers four options for narrowing down a search, including: selecting a specific Dublin Core field to search within, indicating the strictness of the term construction (all, any, exact, and none), searching within a date range, and searching within specific collections. Determining how frequently the advanced search is used is a bit complicated in CONTENTdm. The URL that results from an advanced search query can look the same as a query built from general searches and facet selection. And once a query search is built, that URL can be embedded anywhere and clicked on by users who may or may not know the exact construction of the search. Therefore, URL parameters that match advanced search options cannot be construed as having originated from the advanced search nor that the user intended the search to be performed in that exact way. However, to access the advanced search option when starting a new query, the main URL for the advanced search page is registered in the web logs. In USU's instance of CONTENTDM, the URL looks like: <https://digital.lib.usu.edu/digital/search/advanced>. As noted in Table 1, this URL was viewed only three times over the course of the research project, representing 0.1% of all pageviews. This indicates that advanced searching was not a common tool that users employed.

1.4 Are search terms user-generated or the product of clickable search queries?


USU's CONTENTdm instance also offers many of what the research team deemed "clickable searches" in the form of links. Typically, clickable searches are pre-built search queries that are created in two ways: 1) by library personnel to highlight specific parts of the collection or provide easy access to groups of items or items of note and 2) by the CONTENTdm system for fields designated as a controlled

vocabularies or for facets. As these links proved to be a vital and popular source of navigation in USU's Digital History Collections database, it is important to understand the details of how they are created.

Landing Page Links (personnel-built)

Clickable search links initiated by library personnel are created by conducting a search to narrow down a results list to a very specific group of items. Once this is accomplished, library personnel copy the URL with the search parameters and embed it as a hyperlink. These pre-built search links are typically found on the landing page of a collection. (See Fig. 1) Throughout the article, they will be referred to as Landing Page Links.

Cache Valley Refugee Oral History Project



Cache Valley, Utah is the home of Burmese Muslim, Karen, and Eritrean refugees. Documenting and preserving their stories is an important goal of Utah State University's Fife Folklore Archives (FFA). In May 2015, USU's FFA and Folklore Program, with help from the Karen community, hosted a Library of Congress Field School for Cultural Documentation: "Voices: Refugees in Cache Valley." Field school students worked to document Cache Valley's recent refugee communities. This collection, Folk Coll 58: Cache Valley Refugee Oral History Project, houses the physical recordings and associated materials of the interviews which took place during this field school.

You can explore this collection in two ways:

View the Digital Exhibits:

- [Voices: Burmese Muslims in Cache Valley, Utah](#)
- [Voices: Eritrean Refugees in Cache Valley, Utah](#)
- [Voices: Karen Refugees in Cache Valley, Utah](#)

Or view the individual audio, transcript, and photographs for our interviews:

Burmese Muslim Refugees:

- [Har Be Bar](#)
- [Ka Mar](#)
- [Ka Sin](#)
- [Kaung Lay](#)
- [Sa Jan](#)
- [Ya He Ma](#)

Eritrean Refugees:

- [Afeworki Woldemichael](#)
- [Berhane Debesai Abraha](#)
- [Kahsay Berhe Gebremedhin](#)
- [Seltene Gebreselasie Gebretinsa](#)

Karen Refugees:

- [Aye Win](#)
- [Eh Htoo](#)
- [Kyaw Eh](#)
- [Pyo Nwe](#)
- [Snay Tun](#)
- [Tun Lay](#)

Figure 1. Cache Valley Refugees Oral History Project digital collection landing page demonstrating pre-built search links created by library personnel.

If one were to click on the link for “Afeworki Woldemichael,” as seen in Fig. 1, to view all of the digital objects that pertained to their oral history, the link would open up the following search URL:

<https://digital.lib.usu.edu/digital/collection/p16944coll14/search/searchterm/Afeworki%20Woldemichael/field/creato/mode/all/conn/and/order/nosort> . This link searches the Cache Valley Refugee Oral History Project digital collection for the text string “Afeworki Woldemichael” in the Creator field.

Content in these links may or may not reflect the exact wording, capitalization, or construction of system-built links such as record page links or facet links. In this case, the individual’s name is the search term and the specific field is “Creator” but as will be evident by the next example, it does not exactly match the term supplied in the creator field.

Record Page Links (system-built)

Record pages in CONTENTdm include the digital object at the beginning of the page, followed by a transcript and metadata. Within the metadata, fields set as controlled vocabulary fields (instead of simple text fields) are hyperlinked. These appear as the blue text in Fig 2. When patrons click on the hyperlinked text, CONTENTdm produces a search result list that looks for the same text in the same field throughout the digital collection. Within the Creator field, the hyperlink URL is similar to the search that was found the landing page, with the addition of the birth year of the creator, which was supplied as a text string by the Metadata Librarian :

<https://digital.lib.usu.edu/digital/collection/p16944coll14/search/searchterm/Afeworki%20Woldemichael%201979-/field/creato/mode/exact/conn/and> . This link searches the Cache Valley Refugee Oral History Project digital collection for the text string “Afeworki Woldemichael, 1979-” in the Creator field.

▼ Item Description

Title	Afeworki Woldemichael's portrait May 17, 2015
To read the transcript	http://digital.lib.usu.edu/cdm/ref/collection/p16944coll14/id/94
To listen to the audio	http://digital.lib.usu.edu/cdm/ref/collection/p16944coll14/id/97
Description	Afeworki Woldemichael posing for a photograph. Afeworki Woldemichael is from Eritrea, Africa.
Subject	Refugees Africa Eritrea Ethiopia
Interviewee	Afeworki Woldemichael, 1979-
Interviewer	Warner-Evans, Hilary, 1994-
Other people present	Olsen, Magen, 1986- Berhane Debesai Abraha, 1980-
Photographer	Williams, Heidi 1989-
Translator	Berhane Debesai Abraha 1980-
Language	eng tir
Date	May 17, 2015
Place(s) Immigrated From	Eritrea, http://sws.geonames.org/338010/ Ethiopia, http://sws.geonames.org/337996/ Egypt, http://sws.geonames.org/357994/ Bronx County, New York, United States, http://sws.geonames.org/5110266/ New Haven, New Haven County, Connecticut, United States, http://sws.geonames.org/4839366/ Logan, Cache County, Utah, United States, http://sws.geonames.org/5777544/
Place Interviewed	Logan, Cache County, Utah, United States, http://sws.geonames.org/5777544/
Religious Affiliation	Eritrean Orthodox Christian
Medium	Photographs
Time Periods	21st century 2010-2019
Source	Utah State University, Merrill-Cazier Library, Special Collections & Archives, Cache Valley Refugee Oral History Project, FOLK COLL 58, Item 74
Call Number	Folk Coll 58, Item 74
Collection	Cache Valley Refugee Oral History Project
Collection Inventory	http://archiveswest.orbiscascade.org/ark:/80444/xv67608
Digital Publisher	Digitized by : Utah State University, Merrill-Cazier Library
Rights	Reproduction for publication, exhibition, web display or commercial use is only permissible with the consent of the USU Special Collections & Archives, phone (435) 797-8248.

Figure 2. Cache Valley Refugees Oral History Project digital collection record page demonstrating system-built search links.

Facet Links (system built)

The search results pages in CONTENTdm contain lists of facets along the left-hand side of the webpage. There are usually up to five different facets categories that will filter search results which can vary across collections. The default facets on the main search page are subject, creator, coverage-spatial, type and date. Within each collection, the facet can be specified for the needs of the collection. Most typically, facets for collection-level search are built from subject, creator, coverage-temporal, type, date, or a customized field. For this collection, the creator field is called “Interviewee” and the names included in the creator field, across all digital objects in the collection, are collocated here. Note that the names are all displayed in lower case. This holds true for the URL, as well. When filtering for Afeworki Woldemichael’s digital objects, the URL generated is similar to the record link for their creator name, with the exception of being in lower case.

<https://digital.lib.usu.edu/digital/collection/p16944coll14/search/searchterm/afeworki%20woldemichael%201979-/field/creato/mode/exact/conn/and/order/nosort> This link searches the Cache Valley Refugee Oral History Project digital collection for the text string “afeworki woldemichael, 1979-” in the Creator field.

▼ Interviewee

- kaung lay, 1963- (17)
- gebretinsa, seltene, 1975- (16)
- ka mar, 1988- (16)
- aye win, 1974- (15)
- snay tun, 1994- (15)
- tun lay, 1954- (15)
- berhane debesai abraha, 1980- (14)
- gebremedhin, kabsay berhe, 1971- (14)
- eh htoo, 1983- (13)
- har be bar, 1995- (12)
- ka sin, 1967- (12)
- pyo nwe, 1978- (10)
- afeworki woldemichael, 1979- (9)**
- kyaw eh, 1979- (9)
- ka ma din, 1973- (8)
- ya he ma, 1972 (7)
- pyo nwe, 1978 (2)
- sa jan, 1988- (2)
- ya he ma, 1972- (2)
- snay tun, 1995- (1)
- ya he ma, 197-2 (1)

Show Less

Figure 3. Cache Valley Refugees Oral History Project digital collection facet list for “creator” or interviewee field.

With that background detailed, the answer to the research question “Are search terms user-generated or the product of clickable search queries?” is understandably a bit complex. For each of these “clickable search queries,” the URL recorded in Google Analytics appears the same way regardless of whether they result from clicking on a link on the landing page or the user enters the same search terms and conditions directly. Therefore, the research team cannot fully conclude that all URLs that match landing page links, controlled vocabulary links, or facets links were the result of users clicking on the pre-built links. However, there are some variabilities between the three types of links. Given the unusually high incidence of search queries exactly matching pre-built searches with these variabilities, as noted in Table X, the research team coded the URLs as “likely user-generated” (search terms that do not match any pre-built, clickable search query link available in the system), and “likely clickable query” (search terms in queries that exactly match a pre-built link.)

URL Category	Unique URLs		Pageviews	
	Total Unique URLs	Percent of Total Unique URLs	Total Pageviews	Percent of Total Pageviews
Likely user-generated	127	33.96%	355	47.72%
Likely clickable query	238	63.64%	380	51.08%
Likely combination	9	2.41%	9	1.21%
Total	374		744	

Table 3. Categories of URLs containing search parameters.

As seen in Table 3, users navigated the CONTENTdm website slightly more using clickable queries (personnel-built and system-built links) than by creating search queries directly. This was most pronounced when looking at unique URLs, but still held true when analyzing by the number of pageviews for each category type.

URL Type	URL Category	Unique URLs		Pageviews	
		Total Unique URLs	Percent of Total Unique URLs	Total Pageviews	Percent of Total Pageviews
External link	Likely clickable query	5	1.34%	5	0.67%
Facet link	Likely clickable query	52	13.90%	60	8.06%
Landing Page Link	Likely clickable query	61	16.31%	103	13.84%
Landing Page, then facet applied	Likely clickable query	51	13.64%	99	13.31%

Record page link	Likely clickable query	54	14.44%	94	12.63%
Record page link OR Facet link	Likely clickable query	15	4.01%	19	2.55%
User generated search	Likely user-generated	127	33.96%	355	47.72%
User generated search, then facet applied	Likely combination	9	2.41%	9	1.21%
Total		374		744	

Table 4. Specific URL types and associated categories for URLs containing search parameters.

Table 4 breaks down the URL categories from Table 3 into the specific origins of URLs, looking at how often the clickable queries originated from the landing page, facets, record page, or were from external links. It also includes user generated searches, as a comparison, including when those queries are refined by facets after the initial search. Table 4 demonstrates that the landing page link plays a large role in navigation using clickable queries (comprising just over 27% of the total pageview for URLs with search parameters). Facets, which are unique in the search process because they can refine any of the landing page, record page, or user generated queries, amounted to just over 25% of URLs with search parameters. Record page links came to just over 15% of these same URLs. External links indicate URLs with obvious markers of coming from external websites. For instance, they included URLs that lead to deleted collections (and could not have been searched for using the search parameters). It is assumed by the research team that users clicked on links to old collections maintained on external websites.

Question 2: What search terms were used?

Search URLs, defined as any URL that involved querying the database with any set of parameters, represented 25.3% of all unique URLs and 30% of all pageviews in this project. Search URLs do not necessarily have to include search terms – as they can be generated with a simple query for the contents of an entire collection, often accessed through the ‘Browse this collection’ link on the landing page. URLs with search terms, however, appeared in 374 out of 2,021 URLs, or roughly 18.5% of the unique URLs pulled for this project.

	Count of Unique		Sum of Pageviews	
	Number	Percentage of All URLs	Number	Percentage of All URLs
All URLs	2,021	N/A	3,323	N/A
All search URLs	512	25.3%	997	30.0%
URLs with search terms	374	18.5%	744	22.4%
URLs with general search	138	6.8%	253	7.6%

Table 5. Count of unique and sum of pageviews for search-related URLs.

Search terms appearing in links were extracted and analyzed by the research team. Thirteen categories were identified. The thirteen categories were: address, building, date, filename, first name, format, geographic feature, geographic location, institution, last name, middle name, subject, and title. In many cases, categories of search terms were intermixed, for example first and late names were often used together. Terms were coded independently by at least two different members of the research team and compared for consistency.

Question 2.1: What categories of search terms were used overall?

Of the 374 unique URLs, the “last names” category had the greatest number of unique terms used, while “date” was found in the URLs most frequently and had the largest number of pageviews.

Search Term Category	Unique terms in category		Number of URLs		Sum of pageviews	
	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total
Address	2	1.1%	2	0.3%	3	0.3%
Building	5	2.7%	5	0.8%	7	0.6%
Date	20	10.9%	196	32.0%	427	36.2%
Filename	2	1.1%	3	0.5%	12	1.0%
First Name	27	14.8%	40	6.5%	50	4.2%
Format	16	8.7%	84	13.7%	108	9.1%
Geographic Feature	1	0.5%	1	0.2%	1	0.1%
Geographic Location	15	8.2%	76	12.4%	284	24.0%
Institution	1	0.5%	1	0.2%	1	0.1%
Last Name	41	22.4%	62	10.1%	79	6.7%
Middle Name	3	1.6%	3	0.5%	5	0.4%
Subject	39	21.3%	129	21.0%	185	15.7%
Title	11	6.0%	11	1.8%	19	1.6%
Totals	183	N/A	613	N/A	1181	N/A

Table 6. Overall breakdown of unique terms, number of URLs, and sum of pageviews for search categories.

Question 2.2: What categories of search terms were used most frequently in clickable search queries?

In looking at the breakdown of search categories used in clickable search queries, “subject” and “last name” terms have the highest number of unique terms but similar to the overall categories, the “date” is the largest category in terms of number of URLs and the sum of pageviews. It is also worth noting that the “date,” “filename,” “geographic feature,” and “title” search terms are exclusively found as clickable search queries. This indicates that users are unlikely to type these terms into a search bar but ARE likely to use them when presented in a facet or a clickable landing page or record page link.

Search Term Category	Unique terms in category		Number of URLs		Sum of pageviews	
	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total
Address	0	0.0%	0	0.0%	0	0.0%
Building	2	1.7%	2	0.5%	4	0.5%
Date	20	16.7%	196	47.7%	427	58.6%
Filename	2	1.7%	3	0.7%	12	1.6%
First Name	13	10.8%	13	3.2%	14	1.9%
Format	12	10.0%	47	11.4%	59	8.1%
Geographic Feature	1	0.8%	1	0.2%	1	0.1%
Geographic Location	5	4.2%	5	1.2%	7	1.0%
Institution	0	0.0%	0	0.0%	0	0.0%
Last Name	26	21.7%	26	6.3%	26	3.6%
Middle Name	1	0.8%	1	0.2%	1	0.1%
Subject	27	22.5%	106	25.8%	159	21.8%
Title	11	9.2%	11	2.7%	19	2.6%
Totals	120	N/A	411	N/A	729	N/A

Table 7. Breakdown of unique terms, number of URLs, and sum of pageviews for search categories from clickable search queries.

Question 2.3: What categories of search terms were used most frequently in user generated searches?

In looking solely at the search term categories used in user generated searches, “last name,” “first name,” and “subject” have the greatest number of unique terms. But when looking at the number of URLs (35.1%) and sum of pageviews (61.3%), the most frequently used category is for the “geographic location.” The next most frequent categories are “last name” (17.8% of URLs and 11.7% of pageviews) and “format” (18.3% of URLs and 10.8% of pageviews). This indicates that more unique name and subject terms might be appearing in searching, but that the highly engaged user may be searching for information on a geographic location.

Search Term Category	Unique terms in category		Number of URLs		Sum of pageviews	
	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total
Address	2	3.2%	2	1.0%	3	0.7%
Building	3	4.8%	3	1.5%	3	0.7%
Date	0	0.0%	0	0.0%	0	0.0%
Filename	0	0.0%	0	0.0%	0	0.0%
First Name	14	22.2%	27	13.4%	36	8.0%
Format	4	6.3%	37	18.3%	49	10.8%
Geographic Feature	0	0.0%	0	0.0%	0	0.0%

Geographic Location	10	15.9%	71	35.1%	277	61.3%
Institution	1	1.6%	1	0.5%	1	0.2%
Last Name	15	23.8%	36	17.8%	53	11.7%
Middle Name	2	3.2%	2	1.0%	4	0.9%
Subject	12	19.0%	23	11.4%	26	5.8%
Title	0	0.0%	0	0.0%	0	0.0%
Totals	63	N/A	202	N/A	452	N/A

Table 8. Breakdown of unique terms, number of URLs, and sum of pageviews for search categories from user-generated queries.

Question 3: Where are search terms appearing in the record?

As noted in the first two research questions, we cannot definitively tie search terms to page views or frequency of appearing in searches for users due to the static nature of the CONTENTdm record URLs. In lieu of this, the research team re-ran URLs that included search terms and parameters and scraped all resulting query returns and the metadata from their respective records. In total, 2,933 records were scraped. We then coded each record for where the search term from that specific query occurred in the metadata. Field names were coded by their Dublin Core mappings, with the exception of transcript. In USU's CONTENTdm instance, transcripts can be OCR, machine generated transcription, or hand-generated transcription, but the type or origin of the transcription is not recorded in the repository. This field is not mapped to Dublin Core. As noted in the background section, records typically have a minimum of 21 fields that contain metadata.

Question 3.1: In which fields are search terms appearing overall?

Table 9 outlines the results normalized and collated to Dublin Core fields (not localized field names). Overall, the subject field most commonly held the search terms used by patrons, with over 44% of records containing a matching term in the subject field. Transcripts were the second most common place to find search terms, with 30% of records containing a matching term. Search terms appeared in title, local fields (not mapped to Dublin Core), date, and coverage-temporal fields in between 20 and 22% of records.

Dublin Core Field	Count of Records	Percentage of Records (n=2,933)
Subject	1294	44.12%
Transcript (Not mapped)	880	30.00%
Title	642	21.89%
Local Fields (Not mapped)	615	20.97%
Date	609	20.76%
Coverage-Temporal	605	20.63%
Description	529	18.04%
Source	465	15.85%
Format-Medium	415	14.15%
Coverage-Spatial	330	11.25%
Contributors	192	6.55%

Rights	176	6.00%
Creator	173	5.90%
Relation-IsPartOf	124	4.23%
Relation-IsFormatOf	120	4.09%
Type	91	3.10%
Publisher	71	2.42%
Format	65	2.22%
Relation	24	0.82%
Relation-IsReferencedBy	22	0.75%
Description-Abstract	21	0.72%
Rights-RightsHolder	13	0.44%
Title-alternative	5	0.17%

Table 9. Breakdown of Dublin Core fields where search terms were found, represented by count of records and percentage of total records.

The research team speculates that the impact of subject fields appears to be due to two factors: the use of subject fields in clickable query links and the frequency of personal names in subject fields. Without the ability to tie records directly to search queries, though, we are unable to definitively prove or disprove these theories.

When looking at these results through the lens of field availability – the impact of transcriptions becomes stark. Transcripts were available in 1,007 of the 2,933 records – which is approximately 33% of all the records in this data set. Only 127 of the 1,007 records with transcripts - just 12% - did not match search terms used in the query. In contrast, Subjects were available in 2,539 records (86% of records scraped) but only 1,294 matched search terms or 51% of the records in which it was available.

Question 3.2 How often were search terms only found in one field and which fields were they?

In approximately 48% of results (1,441 records), search terms were found in just one field within the record. This means that without that field, the record would not have appeared in the search results list. When examining how often search terms matched just one field in a record, the impact shifts from subject fields to the transcript, albeit by just a small margin, with transcripts commanding 42.75% of this unique category and the subject field coming it at close second with 41.43%. As demonstrated in Table 10, the coverage-temporal, date, local fields, coverage-spatial, format-medium, description, title, rights, and contributor also made the list of records with search terms found in just one field. However, their impact was significantly less with none breaking more than 7% of records.

Dublin Core Field	Count of Records	Percentage of Records (n=1,441)
Transcript (Not mapped)	616	42.75%
Subject	597	41.43%
Coverage-Temporal	92	6.38%
Date	88	6.11%
Local Fields (Not mapped)	23	1.60%
Coverage-Spatial	13	0.90%
Format-Medium	7	0.49%

Description	3	0.21%
Title	1	0.07%
Rights	1	0.07%
Contributor	1	0.07%

Table 10. Count and percentage of records where search term(s) appear in only one field, broken out by the field in which the term(s) appear.

Discussion

In analyzing how users interact with the digital repository, it was found that most activity (60%) takes place on record pages, with another 30% taking place on search pages. Landing and home pages only account for 9.3% of pageviews and advanced searching only commanded a tenth of 1%. The prominence of record page pageviews over search page pageviews indicate that most users are likely coming directly from a browser search or a direct link. Although the research team could not fully construct the search path, analysis of URLs demonstrated that 88.2% of pageview for record pages were for URLs that never appeared in a search, further highlighting the likelihood of this direct path into the record page for the digital object.

However, when looking at the URLs that include some kind of search term, it was found that “clickable searches” or URLs that are likely the result of a pre-built search query included by practitioners on a record, search, home, or landing pages were slightly more likely to occur than actual user searches. This finding demonstrated that clickable searches on any page (record, search, home, or landing) were important means of navigation within the repository.

In analyzing what and how search terms were utilized in the search process, the research team found that prominent categories of search terms were “date,” “geographic location,” and “subject.” These were often the result of faceting techniques on the part of the user to hone and modify results. This supports the findings of Bogaard et al. (2019) for important fields utilized in the faceting process. For dates, the USU Libraries’ metadata practice of including a date range in the coverage-temporal field (such as 1990-1999 or 20th century) was shown to be utilized by users.

Interestingly, name searches had the largest number of unique terms but only represented a small percentage of the pageviews. Previous research on finding aids conducted by the research team found that name searches were rather high for archival collections and expected the same trend in a digital repository that primarily held digitized archival objects (Daybell et al., 2021). This did not hold true for URLs with search terms in the digital repository. However, this does not necessarily mean that users were not searching names. Given the prominence of direct URLs, the research team cannot definitively say what searches brought users to the record pages but can only say of the search terms actively used within the repository, terms for dates, geographic locations, and subject were utilized by users at a higher rate. Additionally search terms for date, filename, geographic feature, and title were exclusively found as clickable search queries and were not actively keyed in by users. (Note that the filename field is local field used to link related items in a collection.) Therefore, these kinds of search terms may best be presented as controlled vocabulary or actional links.

In examining which fields held search terms most often, the subject field was the most important, with 44.12% of the records appearing in searches due to the subject terms (as noted in the background section, inclusion of at least one Library of Congress Subject Heading is required by institutional practice). The transcript was the second most important field, matching in 30% of the records. However, when considering the availability of fields, the transcript field outperformed the subject field when available. Eighty-eight percent of records with a transcript that appeared in searches matched user query terms

compared to 51% of records with subject fields. The title field, local field (a catch-all category for all customized, non-system, non-mapped fields), date, and coverage-temporal fields also prominently matched with between 20-21% of records containing search terms because of the field content. When looking just at exclusive fields – or fields in which the user’s search terms ONLY matched that field and none other in the record – the transcript field and subject fields were the most important fields in the record. For USU Libraries, the subject field is mapped to Dublin Core and is harvested by aggregators – but the transcript field is not. This indicates that there might be some missing opportunities for users when searching aggregated spaces such as the Mountain West Digital Library (MWDL) or the Digital Public Library of America (DPLA). The high rate of search term occurrence in the transcript field indicates it would be worth opening discussions with content aggregators to explore ways to utilize this content.

When looking at the larger ecosystem of search query effectiveness in library and archival metadata schemas, such as in MARC and EAD, some intriguing discoverability trends arise. Previous research from USU Libraries found that content description fields such as a table of contents or abstract notes in MARC or inventory descriptions in EAD metadata lead to increased discoverability and connection with search terms that users apply in their query process. (Woolcott et al., 2021; Daybell et al., 2022) We would have anticipated that the description field in Dublin Core would perform similarly, but it only matched search terms for 18% of records appearing in the search results. We hypothesize that the subject field played a prominent role because this data set was limited to search queries occurring within the system only, and facet-able or linked fields with controlled vocabularies showed high usage for navigation. But we would still anticipate that the description fields would have performed better. The performance of the transcription field over the description field is particularly perplexing. Again, this is limited only to the search queries from within the system and does not include the much more prominent searches from browsers going directly into records pages. It is an area that the research team would like to explore further, including doing a text analysis.

Considering definitive research results, specific metadata strategies can be continued or implemented to help increase discoverability of resources within CONTENTdm. For example, wherever possible, the use of clickable searches will continue to be incorporated into landing pages and metadata fields, which will group similar items together and cross reference between related collections. Increasing the accuracy and granularity within the most used metadata fields should improve user experience as well. Date ranges (as controlled vocabularies), geographic locations, and subjects can be identified, further refined, and better approximated where possible through consultation with collection donors and curators, crowdsourcing activities, or further research.

Possibilities for further research are promising. In collaboration with other institutions, different digital collection platforms could be examined similarly, to compare discoverability, metadata field usage, the usefulness of system capabilities, or user experience to CONTENTdm. Engagement with metadata across users could also be gauged by duplicating a limited set of collections, with each collection being described differently, for example, the same collection could be published twice, one with descriptive emphasis on expanded description and subject fields, and the other with limited description and subject but emphasis placed on transcriptions. Then, usage data could be gathered to compare the discoverability of each version of the collections.

Other research areas dealing with user engagement can be explored as well. Exploring the efficacy of OCR (or alternative computer-assisted transcription methods) in comparison to transcriptions generated by humans could yield insights into improving the ability of transcription fields to connect with user queries. Collection usage data can also be used similarly to the research conducted by Biswas and Marchesoni (2016) which examined the value of participating in metadata harvesting into repositories like the DPLA.

Finally, additional insights may also be gleaned from examining a larger dataset, with information gathered over a longer period to track user search behavior trends and changes in usage patterns over time. This longitudinal look at user engagement could show the effectiveness (or lack of) for newly adopted metadata strategies. It may also show the strengths and weaknesses of the digital asset management system more clearly.

Conclusion

Web log analysis, web scraping, and data coding of user search terms helped demonstrate how users generally navigate within CONTENTdm, commonly using clickable queries such as controlled vocabularies, facets, and links built by metadata creators to simplify searching for related resources. Additionally, users filter with dates and, in particular, date ranges presented as a controlled vocabulary in a facet are heavily used. When searching with geographic terms, users often present a deeper engagement pattern. Most significantly, though, results of data analysis showed that the transcript and subject fields were the most likely fields to match user search terms. Transcript fields, in particular, were promising, as 88% of records containing this field matched detectable search queries and also comprised 42% of the instances in which only one field (in this case, the transcript field) contained a search term.

The methodology used for this research can be replicated by other institutions using CONTENTdm or similar methodology can be applied within other digital asset management systems to show user search behavior and give insight into ways metadata is utilized in the search process. Gathering user-search data is effective for making evidence-based decisions regarding metadata best practices and determining possible ways to increase the discoverability of library resources.

Acknowledgements

The authors wish to thank our colleague, Kurt Meyer (Cataloger, Utah State University), for his support during this project.

Disclosure Statement

The authors report there are no competing interests to declare.

References

- Biswas, P., Marchesoni, J. (2016). Analyzing Digital Collections Entrances: What Gets Used and Why It Matters. *Information Technology and Libraries*, 35(4), 19-34. <https://doi.org/10.6017/ital.v35i4.9446>
- Bogaard, T., Hollink, L., Wielemaker, J., van Ossenbruggen, J. and Hardman, L. (2019). Metadata Categorization for Identifying Search Patterns in a Digital Library, *Journal of Documentation*, 75(2), 270-286. <https://doi.org/10.1108/JD-06-2018-0087>
- Daybell, P., Payant, A., Woolcott, E., Skeen, B., Arnljots, A. M., & Meyer, K. (2022). A Tale of Two Levels: Analyzing the Discoverability and Impact of Item-Level Description in EAD Finding Aids. *Journal of Archival Organization*, 18(3-4), 101-129. <https://doi.org/10.1080/15332748.2022.2089321>

- Greene, M., & Meissner, D. (2005). More Product, Less Process: Revamping Traditional Archival Processing. *The American Archivist*, 68(2), 208-263. <https://doi.org/10.17723/aarc.68.2.c741823776k65863>
- Gvianishvili, G., Šimko, T., Caffaro, J., Rajman, M., Le Meur, J. Y., Kaplun, S., ... Marian, L. (2010). Capturing and Analyzing User Behavior in Large Digital Libraries, *3rd Workshop on Very Large Digital Libraries, Glasgow, UK, 10 Sep 2010*, 24-31.
- Han, H. J., Joo, S., & Wolfram, D. (2014). Using Transaction Logs to Better Understand User Search Session Patterns in an Image-Based Digital Library. *Journal of the Korean BIBLIA Society for library and Information Science*, 25(1), 19-37. <https://doi.org/10.14699/kbiblia.2014.25.1.019>
- Han, H., Wolfram, D. (2016). An Exploration of Search Session Patterns in an Image-Based Digital Library. *Journal of Information Science*, 42(4), 477-491. <https://doi.org/10.1177/0165551515598952>
- Hanrath, S., & Radio, E. (2017). User Search Terms and Controlled Subject Vocabularies in an Institutional Repository. *Library Hi Tech*. Vol. 35 No. 3, 360-367. <https://doi.org/10.1108/LHT-11-2016-0133>
- Mountain West Digital Library Metadata Application Profile Task Force. (2019). MWDL Metadata Application Profile - version 3. https://github.com/mountainwestdl/mwdl-map/blob/master/MWDL_MetadataApplicationProfile_v3_2019.pdf
- Payant, A., Shaw, M., Arnljots, A., & Woolcott, L. (2020). Utah State University - University Libraries Metadata Application Profile for CONTENTdm Digital Collections, version 2.0. *Digital Commons @ USU*. Available at https://digitalcommons.usu.edu/lib_pubs/308/
- Therrell, G. (2019). More Product, More Process: Metadata in Digital Image Collections, *Digital Library Perspectives*, Vol. 35 No. 1, 2-14. <https://doi.org/10.1108/DLP-06-2018-0018>
- White, P. B., & Radio, E. (2020). Mapping Search Queries to Metadata Fields in a Geoblacklight Repository. *Journal of Library Metadata*, 20(4), 219-235. <https://doi.org/10.1080/19386389.2020.1915459>
- Woolcott, L., Payant, A., Skeen, B., & Daybell, P. (2021). Missing the MARC: Utilization of MARC fields in the search process. *Cataloging & Classification Quarterly*, 59(1), 28-52. <https://doi.org/10.1080/01639374.2021.1881010>
- Zavalina, O., & Vassilieva, E. V. (2014). Understanding the Information Needs of Large-Scale Digital Library Users. *Library Resources & Technical Services*, 58(2), 84-99. <https://doi.org/10.5860/lrts.58n2.84>