7-4-2022

# A Tale of Two Levels: Analyzing the Discoverability and Impact of Item-Level Description in EAD Finding Aids

Paul Daybell
*Utah State University*

Andrea Payant
*Utah State University*

Liz Woolcott
*Utah State University*

Becky Skeen
*Utah State University*

Anna-Maria Arnljots
*Utah State University*

Kurt Alan Meyer
*Utah State University*

# A Tale of Two Levels: Analyzing the Discoverability and Impact of Item-Level Description in EAD Finding Aids

Paul Daybell[a]*, Andrea Payant[a], Liz Woolcott[a], Becky Skeen[a], Anna-Maria Arnljots[a] & Kurt Meyer[a]

[a]*University Libraries, Utah State University, Logan, Utah, USA*

Paul Daybell, paul.daybell@usu.edu, Phone: 435-797-2666

Paul Daybell is the Archival Cataloging Librarian for Utah State University Libraries His research interests include user-search behavior and discoverability of archival resources and technical services outreach. https://orcid.org/**0000-0001-9830-5870**

Andrea Payant, Metadata Librarian for Utah State University since 2017, manages the creation and maintenance of Digital History Collections metadata. She earned an MLIS degree from San Jose State University and has been working in the cataloging and metadata field since 2008. Her research interests include metadata quality benchmarks, metadata workflow efficiencies, and user-search behavior. https://orcid.org/**0000-0001-9873-1538**

Liz Woolcott is the Head of Cataloging and Metadata Services for Utah State University Libraries and the co-founder of the Library Workflow Exchange. Her research interests include digital content reuse and metadata standards assessment. https://orcid.org/**0000-0002-6017-1392 ; Twitter handle: @lizwoolcott**

Becky Skeen is the Special Collections Cataloging Librarian for Utah State University Libraries. Her research interests include outreach for technical services, assessment of special collections cataloging training and professional development, assessment methods for cataloging and metadata services, and optimizing cataloging records based on user search behaviors. https://orcid.org/**0000-0002-0215-6936**

Anna-Maria Arnljots is a Cataloging and Metadata Services Assistant at Utah State University Libraries. She holds an MA in American Studies with an emphasis in Folklore (2000) and MS in Technical Writing (2013), both from Utah State University. Her research interests include crowdsourcing methods, user-search behavior, and ethical metadata practices.

Kurt Meyer is a Cataloger at Utah State University Libraries. He holds master's degrees in English (University of Wyoming) and Library Science (University of Iowa). He previously worked in cataloging and cataloging instruction at Valley City State University.

# A Tale of Two Levels: Analyzing the Discoverability and Impact of Item-Level Description in EAD Finding Aids

As part of a multi-faceted research project examining user engagement with various types of descriptive metadata, Utah State University Libraries Cataloging and Metadata Services unit (CMS) investigated the discoverability of local Encoded Archival Description (EAD) finding aids. The research team put two versions of the same finding aid online with one described at the file (box or folder) level and the other at the item-level. Over a year later, the team pulled the analytics for each guide and assessed which descriptive level was most frequently accessed. The research team also looked at the type of search terms patrons utilized and where in the finding aid they were located. Usage data shows that personal names are the most common type of search term, search terms are most commonly found in the Collection Inventory, and that the availability of item-level description improves discovery by an average of 6,100% over file-level descriptions.

Keywords: Encoded Archival Description, discoverability, archival description, EAD, finding aids, user search behavior

## Introduction

Utah State University Libraries support a multi-campus research institution serving a population of 28,000 students throughout the state of Utah. Within USU Libraries, the Cataloging and Metadata Services (CMS) unit is grouped into a shared department with Special Collections and Archives (SCA) and Digital Initiatives (DI). These three units are charged with working together to make the special or unique holdings within the library open and discoverable to the campus community and beyond.

Currently, EAD finding aid creation and remediation of legacy guides takes place in both the SCA unit and the CMS unit, depending on the archival material category or the type of work required. Finding aids created by the CMS unit are first generated in XML, with the <dsc> populated from an Excel spreadsheet using a mail merge process described in Woolcott et al.[1] Once the finding aid is created, it is loaded into two separate locations: ArchivesSpace (which is only open to staff for viewing) and Archives West, a consortium to which Utah State University Libraries contributes its finding aids for public discovery.

Starting in 2019, the CMS unit embarked on a multi-phase research project to analyze how the metadata schemas and standards employed by the library interacted with user search patterns. This long-term research project analyzes MARC records, EAD finding aids, and Dublin Core records to determine how users are searching for content and where the search terms they apply are found in the records. As an outcome of this long-term project, the research team will attempt to establish a sense of where the efforts of the unit in creating records in each schema or standard can best support discoverability for users. For instance, the research team examined 13,000 MARC records displayed to users in the search process to determine that search terms are most often found in the 245 (Title), 505 (Formatted Content Notes), and 520 (Summary, etc.) fields and that the

505, in particular, played an under-appreciated role in discovery. Additionally, the research project found that authorized name headings were more frequently used than authorized subject headings.[2] As a result of this research, the unit shifted their work to emphasize the addition or enhancement of the abstract and table of contents fields, as well as name authority work to increase the discoverability for library holdings. This article will discuss a similar process for EAD finding aids. In addition, however, it will also attempt to identify if the level to which a collection is described in a finding aid impacts its discoverability. The following outlines the questions addressed in this article:

(1) What type of search terms are most commonly used by patrons and where are they found in the finding aid?

(2) Is there a measurable difference in discoverability between finding aids described to the item-level and finding aids described to a box or folder level?

(3) How do search parameters impact discoverability for file-level description vs. item-level description?

**Literature Review**

Diving into the literature of EAD discoverability, a wide array of research has been conducted which is congruent with various aspects of evaluating EAD and user engagement. The most relevant work done related to this study deals with levels of archival processing and description as well as usage data analytics for finding aids.

Examining how users search, what types of search terms they employ, and where those search terms can be found in a record yields intriguing avenues to explore. Daniels and Yakel[3] reported interesting results regarding user search behavior. They found that users employed traditional search strategies like using Boolean operators and narrowing searches, but that they also adapted searches, for example, using CTRL+F to search copious amounts of text for keywords. Using archival terminology and selecting search terms were identified among impediments to a successful search. Participants in this study were shown to be able to identify keywords but struggled coming up with different terms to use when the original search terms failed to retrieve any results. The authors also found that users benefitted from prior knowledge of controlled vocabularies, knowledge of the topic, and being able to differentiate between different elements of a finding aid.

Yang[4] looked at user search terms and where they were found in Dublin Core and VRA Core records. Yang's results demonstrated that search terms corresponded with title and description fields most frequently. The subject element also played a key role in term matching but could be considered overall less significant because the keywords found in subject fields were also likely to be listed elsewhere in the record. Dublin Core and

VRA Core records describe objects at an item level, instead of a collection level, so Yang's findings on search terms cannot be applied in a direct comparison with EAD but they do point to the potential for search terms being found in component level titles and scope and content notes. Yang also recommended that metadata creators consider grouping fields into functionalities, like discoverability, and designing internal or best practices for those fields to maximize that functionality. This framework is useful for thinking about how descriptive metadata practices can focus work for long-term or user-centered utility.

Bronn, Proffitt, and Washburn[5] analyzed EAD finding aids aggregated from OCLC's ArchiveGrid discovery system, and then replicated their study in 2021 using finding aid data from twelve National Finding Aid Network (NAFAN) partners.[6] These studies included an examination of how well these finding aids support online discovery. Their analysis focused on the EAD schema and its elements specifically. Their work concluded that while the metadata standard met requirements needed for effective discovery, there is a disconnect between the schema itself and established best practices implemented among archives professionals that may affect discoverability. The authors include possible ways for improving EAD, particularly to populate the key elements (date, extent, collection title, subject, material type, names, scope, and biographical notes, and abstract) and to increase the functionality and discoverability of container lists. They also recommend finding ways to incorporate geolocation functionality into the structure of EAD.

When looking at research on the difference in discoverability between item-level and box or folder level description, it is necessary to set the foundation by discussing the most influential articles on archival processing to date, written by Greene and Meissner.[7] The authors discuss the importance of 'more product, less process' (MPLP) for archival collections to increase the rate of backlog processing. One of the methods they propose to increase the rate of getting collections into the hands of patrons is to decrease item-level processing and to be selective in identifying collections that deserve more detailed description.

Meissner and Greene also wrote a follow-up article that addresses the impact of MPLP.[8] The article reiterates the importance of utilizing existing resources at their greatest efficiency and efficacy to eliminate paper backlogs. They emphasize that MPLP is an effective tool for archivists to help achieve this goal. The authors also discuss item-level description but do so by bouncing back at critics who support item-level description for digitized archival material. They give two reasons for rejecting the need for item-level processing. First, they point out that item-level processing is not required for digitization. And second, they argue that it is a false assumption to insist that item-level description is necessary or even the best option for digital objects and that nothing other than preoccupation with individual items prevents archivists from digitizing at a folder or collection level.

However, the archives community has struggled with assessing the impact that MPLP practices have on user discoverability.[9] Tia Edmunson-Morton conducted a study of user data for MPLP processed collections and concluded that user search behavior is complex.[10] Her study concluded that researchers did not necessarily want more description, but preferred description with sufficient context, skim-able lists, and greater uniformity across finding aids. Crowe and Spilman conducted broad surveys of

American archivists and concluded that most research staff believed that MPLP has improved discoverability and user access to archival collections.[11]

Parilla, Morgan, and Fidler[12] focus on the critical needs of certain research fields that necessitate greater levels of description, particularly to provide access to materials such as field notes. The focus of their research illustrates the need to revamp the systemic workflows to meet user research needs and possible methods for improving online access to archival materials. The authors highlight three cases at three institutions with the need to make research data discoverable online. They discuss the use of MODS records, EAD, and ArchivesSpace as methods for providing access to materials. They concluded that standards used to best describe individual items, like MARC or an equivalent, should continue to be used, at least until libraries and archives move beyond MARC into linked data systems like BIBFRAME.

Aleman, Christner, and McGee compiled an extensive literature review that informs readers about methods for improving archival discoverability.[13] The authors highlight works that study technical tools used by archivists for archival description, techniques that can be implemented to improve description, and user interaction with finding aids. They examine numerous avenues to mix old-school vs. new-school ways of discovering archival information and collections, ultimately suggesting that a compromise may be found in using crowdsourcing models and other user-centric models to bridge the gap.

Higgins, Hilton, and Dafis[14] challenge current archival practices and suggest undertaking user studies of alternative methods for providing contextual understanding for archival collections without pre-defined arrangement and minimal description. The authors posit opportunities to describe collections and streamline the process of providing access to digitized archival materials. Opportunities include user generated arrangement and description, user tagging, and linking to existing resources.

In their article, Zhang and Mauney[15] study archives that attempt to develop a connection between archival description and digitized items. The importance of granularity in description of digital items and the complexity of providing archival context is acknowledged by the authors. They outline several strategies that have been developed by different institutions to describe digitized archival collections and make digital objects accessible online. Strategies include three models of representation: embedded, segregated, and parallel. The embedded model is used to describe collection content emphasizing archival context, in which the primary access point is a finding aid and digital counterparts are built in the hierarchical structure of the finding aid. The segregated model is used to describe collection content emphasizing digital content, in which multi-faceted metadata is used as the primary access point. The parallel model is used to describe collection content that emphasizes archival context and digital content, which combines the first two models to supply access points for discovery.

In the article More Product, More Process: Metadata in Digital Image Collections,[16] Therrell examines levels of description and resource retrieval. The author compares user behavior and the discoverability between digital images on an institutional website (less descriptive) and Flickr (more descriptive) in the context of the MPLP theory. Using a methodology of vocabulary analysis and retrieval testing, outcomes showed that retrieval and findability is hindered by less descriptive levels of metadata for digital objects.

## Methodology

With complexities regarding processing times and impacts on discoverability in mind, the research team brainstormed ways to empirically assess what impact the level of description had on discoverability. In so doing, the team settled on creating two identical versions of the same EAD finding aid with the only variation being in the <dsc> description, with one version described to the item and the other to the box or folder level. Both versions of the EAD finding aid would be posted online at the same time and left untouched for a minimum of 1 year. After that time, the web analytics for each of the finding aids would be pulled and analyzed for usage patterns. The research team felt that this method of measuring the two descriptive levels side by side would provide the most concrete evidence of whether the descriptive level impacted discoverability.

The collections chosen for this project came from the University Archives and the Fife Folklore Archives – two distinct sub-archives within the SCA unit. Curators in these units were asked to select 2-3 collections for description, with the caveat that the collection could have no existing online presence, in order to not skew results. The collections selected were:

*University Archives*

- **Utah State University College Journal Index (USU 10.02:27)**: Collection of 16 boxes of newspaper clippings covering the University from 1890 to 1954
- **Utah State University Football Programs (USU 16.1/2:55)**: Collection of 16 boxes of programs for football games (including team rosters and biographies of players/coaches) from 1904-2012
- **Utah State University Men's Basketball Programs (USU 16.1/3:55)**: Collection of 11 boxes of programs for basketball games (including team rosters) from 1948-1991

*Fife Folklore Archives*

- **Adam's Elementary Valentine's Tea Fieldwork (FOLK COLL 69)**: Collection of three boxes of interviews, slides, and cassette tapes documenting a Valentine's Tea at a local elementary school from 1995-2001
- **Bear River Heritage Barn Survey (FOLK COLL 29a)**: Collection of 14 boxes of interviews, photographs, and fieldnotes documenting historic barns in the Bear River Heritage Area, from 2002-2016

For the sake of simplifying discussion, this article will break down the sections of a finding aid into three categories: **Frontmatter**, **Collection Inventory**, and **Control Access**. The **Frontmatter** includes the description of the collection as a whole and typically involves the following tags from the <archdesc> such as: <accessrestrict>, <acqinfo>, <arrangement>, <did><unitid>, <scopecontent>, <did><origination><persname>, <did><unitdate>, <bioghist>, <did><langmaterial>, <prefercite>, <processinfo>, <did><physdesc><extent>, <relatedmaterial>, <did><repository><corpname>, <userestrict>, <did><abstract>, and <did><unittitle>.

Please note that this is different from the <frontmatter> tag itself. This categorical name originates from internal documentation. This section has no formal label in the public display. The **Collection Inventory** includes the descriptions of the individual boxes, folders, or items within the collection and involves the following tags in the <archdesc><dsc><c0X> (where the "x" is supplied with the numerical value for the component level): <did><scopecontent>, <did><unittitle>, <did><unitdate>, <did><container>, <did><unitid>, and <physdesc><extent>. In the public display of the finding aid, this section is called "Detailed Description of Collection". The **Control Access** section of the finding aid includes the authorized headings associated with the collection as a whole and involves the following tags in the <archdesc><controlaccess>: <subject>, <geogname>, and <persname>. In the public display of the finding aid, this section is labelled "Names and Subjects."

For each collection, two versions of an EAD finding aid were produced. Each pair of files was identical in every element including the Frontmatter and Control Access sections. The only difference between each version of the finding aid was the extent to which the Collection Inventory was described, with the most specific component description listed at either the <c level="file">or the <c level="item"> attribute. Therefore, one finding aid with the lowest <c level="file">element would correspond to a description of the physical box or folders in the collection and the other finding aid with a <c level="item"> attribute corresponded to a description of items in the collection. *Image 1* demonstrates how a file-level description appeared to patrons, with *Image 2* showing the underlining XML.

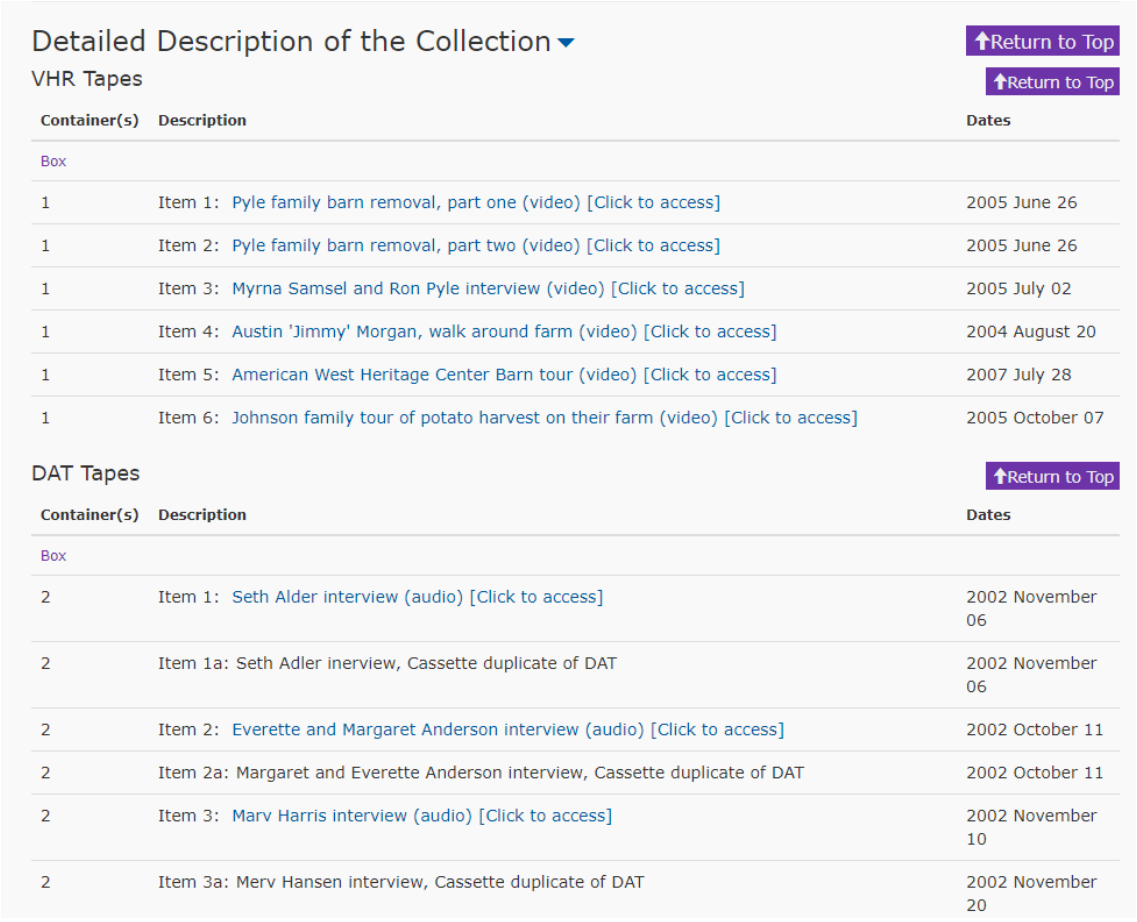
Image 1. Patron view of file-level description


Image 2. XML representation of file-level description


Note that in *Image 1 – Patron view of a file-level description*, the finding aid lists each box in the collection with the box title represented as a link to access the digitized video and audio files for that box. The link for box or folder level description corresponds to a tag search in CONTENTdm for all items tagged with that box or folder id. *Image 2 – XML representation of a file-level description* show that <c01> tags in this example are given a series level attribute and described with a <unittitle>. <c02> levels are given a <c02 level="file">attribute and described with a <container>, <unittitle>, <daogrp> (where applicable), and <unitdate> tag.

*Image 3* below shows how an item-level description appeared to patrons. *Image 4* following shows the underlining XML.

Image 3. Patron view of item-level description



## Detailed Description of the Collection

Return to Top

### VHR Tapes

Return to Top

| Container(s) | Description | Dates |
|---|---|---|
| Box | | |
| 1 | Item 1: Pyle family barn removal, part one (video) [Click to access] | 2005 June 26 |
| 1 | Item 2: Pyle family barn removal, part two (video) [Click to access] | 2005 June 26 |
| 1 | Item 3: Myrna Samsel and Ron Pyle interview (video) [Click to access] | 2005 July 02 |
| 1 | Item 4: Austin 'Jimmy' Morgan, walk around farm (video) [Click to access] | 2004 August 20 |
| 1 | Item 5: American West Heritage Center Barn tour (video) [Click to access] | 2007 July 28 |
| 1 | Item 6: Johnson family tour of potato harvest on their farm (video) [Click to access] | 2005 October 07 |

### DAT Tapes

Return to Top

| Container(s) | Description | Dates |
|---|---|---|
| Box | | |
| 2 | Item 1: Seth Alder interview (audio) [Click to access] | 2002 November 06 |
| 2 | Item 1a: Seth Adler inerview, Cassette duplicate of DAT | 2002 November 06 |
| 2 | Item 2: Everette and Margaret Anderson interview (audio) [Click to access] | 2002 October 11 |
| 2 | Item 2a: Margaret and Everette Anderson interview, Cassette duplicate of DAT | 2002 October 11 |
| 2 | Item 3: Marv Harris interview (audio) [Click to access] | 2002 November 10 |
| 2 | Item 3a: Merv Hansen interview, Cassette duplicate of DAT | 2002 November 20 |

Image 4. XML representation of item-level description

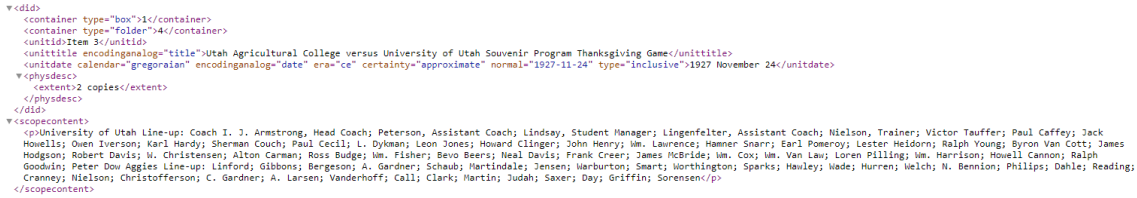
```
<dsc type="in-depth">
 <c01 level="series">
  <did>
   <unittitle encodinganalog="title">VHR Tapes</unittitle>
  </did>
  <c02 level="item">
   <did>
    <container type="box">1</container>
    <unitid>Item 1</unitid>
    <unittitle encodinganalog="title">Pyle family barn removal, part one (video)</unittitle>
    <daogrp>
     <resource xlink:label="start">Pyle family barn removal, part one (video) [Click to access]</resource>
     <daoloc xlink:label="MovingImage" xlink:href="http://n2t.net/ark:/85142/t4121216191325" xlink:role="video/mp4"/>
     <arc xlink:from="start" xlink:to="MovingImage" xlink:show="new" xlink:actuate="onRequest"/>
    </daogrp>
    <unitdate encodinganalog="date" normal="2005-06-26" era="ce" certainty="approximate" calendar="gregorian">2005 June 26</unitdate>
   </did>
  </c02>
  <c02 level="item">
   <did>
    <container type="box">1</container>
    <unitid>Item 2</unitid>
    <unittitle encodinganalog="title">Pyle family barn removal, part two (video)</unittitle>
    <daogrp>
     <resource xlink:label="start">Pyle family barn removal, part two (video) [Click to access]</resource>
     <daoloc xlink:label="MovingImage" xlink:href="http://n2t.net/ark:/85142/t4821216191325" xlink:role="video/mp4"/>
     <arc xlink:from="start" xlink:to="MovingImage" xlink:show="new" xlink:actuate="onRequest"/>
    </daogrp>
    <unitdate encodinganalog="date" normal="2005-06-26" era="ce" certainty="approximate" calendar="gregorian">2005 June 26</unitdate>
   </did>
  </c02>
  <c02 level="item">
   <did>
    <container type="box">1</container>
    <unitid>Item 3</unitid>
    <unittitle encodinganalog="title">Myrna Samsel and Ron Pyle interview (video)</unittitle>
    <daogrp>
     <resource xlink:label="start">Myrna Samsel and Ron Pyle interview (video) [Click to access]</resource>
     <daoloc xlink:label="MovingImage" xlink:href="http://n2t.net/ark:/85142/t431216191325" xlink:role="video/mp4"/>
     <arc xlink:from="start" xlink:to="MovingImage" xlink:show="new" xlink:actuate="onRequest"/>
    </daogrp>
    <unitdate encodinganalog="date" normal="2005-07-02" era="ce" certainty="approximate" calendar="gregorian">2005 July 02</unitdate>
   </did>
  </c02>
  <c02 level="item">
   <did>
    <container type="box">1</container>
    <unitid>Item 4</unitid>
    <unittitle encodinganalog="title">Austin 'Jimmy' Morgan, walk around farm (video)</unittitle>
    <daogrp>
     <resource xlink:label="start">Austin 'Jimmy' Morgan, walk around farm (video) [Click to access]</resource>
     <daoloc xlink:label="MovingImage" xlink:href="http://n2t.net/ark:/85142/t41091216191325" xlink:role="video/mp4"/>
     <arc xlink:from="start" xlink:to="MovingImage" xlink:show="new" xlink:actuate="onRequest"/>
    </daogrp>
    <unitdate encodinganalog="date" normal="2004-08-20" era="ce" certainty="approximate" calendar="gregorian">2004 August 20</unitdate>
   </did>
  </c02>
  <c02 level="item">
   <did>
    <container type="box">1</container>
    <unitid>Item 5</unitid>
    <unittitle encodinganalog="title">American West Heritage Center Barn tour (video)</unittitle>
    <daogrp>
     <resource xlink:label="start">American West Heritage Center Barn tour (video) [Click to access]</resource>
     <daoloc xlink:label="MovingImage" xlink:href="http://n2t.net/ark:/85142/t4781216191325" xlink:role="video/mp4"/>
     <arc xlink:from="start" xlink:to="MovingImage" xlink:show="new" xlink:actuate="onRequest"/>
    </daogrp>
    <unitdate encodinganalog="date" normal="2007-07-28" era="ce" certainty="approximate" calendar="gregorian">2007 July 28</unitdate>
   </did>
  </c02>
```

Note that in *Image 3 – Patron view of item-level description*, the finding aid lists each box and the items within that box for the collection. Each item is linked directly to the digitized video and audio files for that box. *Image 4 – XML representation of item-level description* show that <c> levels are the same as the file-level description, with the exception that the <c02> levels are encoded as @item or <c02 level="item">. <c01> levels in this example are given a @series level attribute and described with a <unittitle>. As noted, <c02> levels are encoded as @item and described with a <container>, <unitid>, <unittitle>, <daogrp> (where applicable), and <unitdate>. Two collections, at the direction of the curator, also included the <scopecontent> and <extent> tags. The <unitid>, <scopecontent>, and <extent> are the only additional tags added beyond what was included in the file-level description. For the remainder of this article, <c0X level="file"> will be referred to as "file-level" and <c0X level="item"> will be referred to as "item-level".

The primary difference is in how many <c02> are added to each guide and in the number and specificity of terms applied to the searchable <unittitle> and <scopecontent> text within the <c02>. The finding aids for the Utah State University Football Programs and the Utah State University Men's Basketball Programs are the

two collections that included the <scopecontent> and, where applicable, the <extent> tags. (See *Image 5*). These tags were included after discussions with the curator noted that a sizable number of patron inquiries was for references to individuals who attended or taught at the university previously. The <scopecontent> note, in particular, was used to house names of participants in sporting events for each program.

Image 5. Additional tags for the Utah State University Football programs and the Utah State University Men's Basketball programs




Only two collections in the project included the <daogrp> tag for accompanying digitized or born-digital collections: the Bear River Heritage Barn Survey (shown above) and the Adam's Elementary Valentine's Tea Fieldwork Collection. Both collections are from the Fife Folklore Archives, which has a strong digital presence. For digitized or born-digital collections, USU Libraries often creates the <dsc> portion of the finding aids from the Dublin Core metadata instead of the other way around. This process was outlined in detail in Woolcott et al.[1] In cases such as this, digital files were described individually and loaded into CONTENTdm. Once there, items were tagged by the box and folder in which they were organized. This action aggregated all items with the same tag together and the tag itself served as a link that displayed all items in those folders or boxes together. This provided equivalent digital access for finding aids described at the individual item-level or the file/box level. The digital collection went live shortly before the finding aid in order to be able to provide the links for the finding aid.

Once the two finding aids for each collection were ready, they were posted at the same time to the Archives West portal. The research team designated one person to review the online finding aids for issues or problems on the day it was posted. After this point, no research team member was permitted to search for or pull up the finding aids. The goal was to see how much traffic each finding aid was able to attract naturally. However, other library and archival staff were not restricted from accessing the finding aids because they regularly assist patrons with searching the collections. Apart from the curator and the research team, no other person was informed about which collections were chosen for the project. *Table 1* below shows the date each collection went live online, the number of days it was available, the count of unique URLs accessed by patrons, and whether the finding aid linked to a digital collection.

Table 1. Collection days online, count of unique URLs, and links to digital content

| **Collection Name** | **Level of Description** | **Date Live** | **Data Collected** | **Days Online** | **Count of Unique URLs** | **Linked to Digital Collection** |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Utah State University College Journal Index | File | 8/15/2019 | 6/16/2021 | 671 | 2 | No |
| Utah State University College Journal Index | Item | 8/15/2019 | 6/16/2021 | 671 | 10 | No |
| Utah State University Football Programs | File | 11/22/2019 | 6/16/2021 | 572 | 6 | No |
| Utah State University Football Programs | Item | 11/22/2019 | 6/16/2021 | 572 | 97 | No |
| Adams Elementary Valentine's Tea Fieldwork | File | 12/17/2019 | 6/16/2021 | 547 | 0 | Yes |
| Adams Elementary Valentine's Tea Fieldwork | Item | 12/17/2019 | 6/16/2021 | 547 | 0 | Yes |
| Bear River Heritage Barn Survey | File | 1/7/2020 | 6/16/2021 | 526 | 3 | Yes |
| Bear River Heritage Barn Survey | Item | 1/7/2020 | 6/16/2021 | 526 | 6 | Yes |
| Utah State University Men's Basketball Programs | File | 5/5/2020 | 6/16/2021 | 407 | 3 | No |
| Utah State University Men's Basketball Programs | Item | 5/5/2020 | 6/16/2021 | 407 | 19 | No |
| | | | | Total URLs | 146 | |

As noted in *Table 1*, each collection had a different start date because the research team had to develop a descriptive inventory for each collection from scratch. Once the last collection was posted online, the research team let the collections sit for one year. On June 17, 2021, the research team downloaded the traffic data for each version of each collection using Google Analytics starting with the day after the EAD finding aid was posted (to eliminate the pageviews generated when the designated reviewer verified that the content met USU Libraries' standards for posting online) and ending with June 16, 2021, the day before the usage data was collected. This meant that collections were online between 13 and 18 months.

Along with the pageviews, the research team downloaded the URLs accessed each day for each collection from Google Analytics. A typical URL includes the ARK assigned to the finding aid, and if present, information about the search parameters chosen by the user, including search terms and filters. When a URL is simply the ARK for the finding aid, it often indicates that it was found through a browser search. In this case, search terms are not known to the research team. When the URL included search terms or system filters, such as the unique repository symbol, it often, though not always, indicated that the user searched within the Archives West system or came through a filtered search from the USU Libraries homepage. The Archives West consortium hosts and displays the EAD finding aids alongside the finding aids of other consortia members. The Google Analytics for the site only records the search terms used within the Archives West website. Any search terms input into a browser search that ended in a user accessing Archives West are recorded in Google Analytics as simply the URL for the collection, without search parameters included. This typically looked like:

https://archiveswest.orbiscascade.org/ark:/80444/xv901763

Whereas any search within Archives West results in a URL that includes the search parameters of the user. This typically looked like:

https://archiveswest.orbiscascade.org/ark:/80444/xv901763/op=fstyle.aspx?t=k&q=barn+survey&f_repo=US-ula

Note the additional information after the ARK. The portion beginning with "q=" and ending before the ampersand included search terms or the query input by the patron in Archives West. The portion of the URL following "repo=" filtered for just the Utah State University Libraries finding aids.

Once downloaded, the pageviews, dates of access, and the URLs accessed were loaded into an Airtable base. URLs were then coded for the type of URL (static or with search terms) and whether the USU repository was selected. Search terms were extracted from the URLs wherever present and coded for where they were found in the EAD finding aid and the type of search term used. This process was repeated a second time by a different coder to ensure intercoder reliability. In total, 146 unique URLs were accessed among the 10 finding aids, with 140 of the URLs including search parameters. Please note this is different from the number of pageviews each URL received. The 6 URLs with no search parameters were viewed 1,026 times whereas the 140 URLS with search parameters included were viewed 157 times (see *Table 2.*) All analysis for Research Question #1 and #3 will be based on the 140 unique URLs that included search parameters.

Table 2. Collection days online, count of unique URLs, and links to digital content

| **URL Type** | **Number of Unique Occurrences** | **Pageviews** |
|---|---|---|
| No search parameters | 6 | 1026 |
| Search parameters included | 140 | 157 |
| **Total** | **146** | **1183** |

## Analysis

### *Question 1: What type of search terms are most commonly used by patrons and where are they found in the finding aid?*

#### *1.1 What types of search terms were used?*

The search terms used by patrons were primarily personal names. As noted in *Table 3* below, last names appeared in 77.86% of URLs with search terms while first names appeared in 57.14% and middle names appeared in 5.71% of URLs. The next highest category, subject terms, was used in 14.29% of URLs and institutional names were used in 8.57% of URLs. The name of a collection and the name of an organization each appeared in 3.57% of the URLs with search terms. Institutional name referred to the

name of a university such as “Utah State University” or “USU,” while organizational names referred to entities within institutions such as the “Home Economics Club.”

Table 3. Search query type frequency

| **Search Query Type** | **Occurrence in URLs *(n=140)*** | **Percentage of URLs** |
|---|---|---|
| Last Name | 109 | 77.86% |
| First Name | 80 | 57.14% |
| Subject | 20 | 14.29% |
| Institutional Name | 12 | 8.57% |
| Middle Name | 8 | 5.71% |
| Collection Name | 5 | 3.57% |
| Organization Name | 5 | 3.57% |
| Place | 4 | 2.86% |
| Year | 3 | 2.14% |
| Collection Number | 2 | 1.43% |
| Title (Person) | 2 | 1.43% |
| Campus location | 1 | 0.71% |
| Format | 1 | 0.71% |

*1.2 In which sections of the finding aid were search terms most commonly found?*

As noted in the Methodology section, the three major sections in USU Libraries’ finding aids are: Frontmatter, Collection Inventory (labelled as “Detailed Description of Collections” in the finding aid), and the Control Access (labelled as “Names and Subjects” in the finding aid). The Frontmatter section includes all the collection-level information such as title, creator, overall date range, collection number, repository information, instructions on use and access, summaries of the collection, and citation information. The Collection Inventory includes the list of series, boxes, folders, or items along with their dates and any scope or content notes about individual items, folders, or boxes. The Control Access section includes authorized headings for subject and geographic terms, as well as personal names. The Methodology section of this article outlines the specific tags used in each section.

*Table 4* shows the finding aid sections in which search terms were most commonly found. The Collection Inventory section was by far the most pivotal section for matching user search terms, connecting with 95% of the URLs that included search terms. The Frontmatter was a very distant second in terms of matching user search terms, connecting with only 20% of the URLs that included search terms. The Control Access section only connected with 2.9% of URLs that contained search terms. However, it should be noted that only 2 collections (Folklore collections Adams Elementary Valentine’s Tea and Bear River Heritage Area Barn Survey) included controlled access terms because the Folklore Archive and the University Archive had

different policy decisions for minimal descriptions at the time the finding aids were created.

Table 4. Section of finding aid where search terms are found

| Section of Finding Aid | Number of URLs (n=140) | Percentage of Total URLs with Search Terms |
|---|---|---|
| Collection Inventory | 133 | 95.0% |
| Frontmatter | 28 | 20.0% |
| Control Access | 4 | 2.9% |

*1.3 In which tags are search terms most commonly found?*

*Table 5* shows the labelled headings of the finding aid, split among the three major sections of the guides: Frontmatter, Collection Inventory, and Control Access Terms. Headings are shown along with their associated tags. Of the 26 subsections in the finding aid, 16 subsections contained the search terms used by patrons.

In the Frontmatter, these subsections were: Acquisition Information, Collection Number, Content Description, Creator, Historical Note, Preferred Citation, Related Material, Repository, Restrictions on Use, Summary, and Title.

Six sections did not have any search terms present: Access Restrictions, Dates, Quantity, Languages, Arrangement, and Processing Note.

In the Collection Inventory, the subsections in which search terms were found included: Scope and Content, Unit Title, and Unit Date. Container, Unit ID, and Extent did not contain any search terms.

In the Control Access Terms section, the subsections where search terms were found were: Subject Terms and Geographical Names. Personal Names did not match any search terms. Please note that only the Fife Folklore collections included terms in their Control Access Terms section as a policy. The University Archives did not, at the time, have a policy to include terms as a standard practice.

Overall, search terms were overwhelmingly found in the Collection Inventory, labelled as “Detailed Description of Collection” in the public view of the finding aid, with 95% found in this section. This section included tags that are presented without formal labels – such as the <unittitle>, <unitdate>, <scopecontent>, and <extent>. They demonstrate that the addition of the personal names in the <scopecontent> notes in the collection inventory had a significant impact on discoverability of the finding aids. The <scopecontent> notes matched the search term for 81.4% of the URLs that contained search terms, with the <unittitle> matching 22.1% of those URLs. These two tags outpaced all other tags in the Frontmatter and Control Access terms. While the <unittitle> played a relatively prominent role in the discoverability of the finding aid, it was significantly less than the <scopecontent> tag. The <scopecontent> tags in item-

level finding aids were primarily used to record names of participants in sporting events, which indicated that the actual content of the item (beyond the title supplied by the processor) may be important to users. Interestingly enough, the next two most common tags in which search terms were found were the <userestrict> and <prefercite>, found in the Frontmatter. These sections were most closely associated with search terms that used the university's name or acronym.

Table 5. Tags where search terms are found

| **Tag Label*** | **Tag Hierarchy** | **Number of URLs (n=140)** | **Percentage of Total URLs with Search Terms** |
|---|---|---|---|
| ***Frontmatter*** | | 28 | 20.0% |
| **Restrictions on Use** | <archdesc><userestrict> | 24 | 17.1% |
| **Preferred Citation** | <archdesc><prefercite> | 23 | 16.4% |
| **Title** | <archdesc><did><unittitle> | 22 | 15.7% |
| **Summary** | <archdesc><did><abstract> | 20 | 14.3% |
| **Historical Note** | <archdesc><bioghist> | 20 | 14.3% |
| **Acquisitions Information** | <archdesc><acqinfo> | 14 | 10.0% |
| **Repository** | <archdesc><did><repository><corpname> | 9 | 6.4% |
| **Creator** | <archdesc><did><origination><persname> | 6 | 4.3% |
| **Content Description** | <archdesc><scopecontent> | 3 | 2.1% |
| **Collection Number** | <archdesc><did><unitid> | 3 | 2.1% |
| **Related Material** | <archdesc><relatedmaterial> | 3 | 2.1% |
| **Access Restrictions** | <archdesc><accessrestrict> | 0 | 0.0% |
| **Dates** | <archdesc><did><unitdate> | 0 | 0.0% |
| **Quantity** | <archdesc><did><physdesc><extent> | 0 | 0.0% |
| **Languages** | <archdesc><did><langmaterial> | 0 | 0.0% |
| **Arrangement** | <archdesc><arrangement> | 0 | 0.0% |
| **Processing Note** | <archdesc><processinfo> | 0 | 0.0% |
| ***Collection Inventory (labelled "Detailed Description of Collection")*** | | 133 | 95.0% |
| *Scope and Content* | <archdesc><dsc><c0x><did><scopecontent> | 114 | 81.4% |

| | | | |
|---|---|---|---|
| *Unit Title* | <archdesc><dsc><c0x><did><unittitle> | 31 | 22.1% |
| *Unit Date* | <archdesc><dsc><c0x><did><unitdate> | 3 | 2.1% |
| *Container* | <archdesc><dsc><c0x><did><container> | 0 | 0.0% |
| *Unit ID* | <archdesc><dsc><c0x><did><unitid> | 0 | 0.0% |
| *Extent* | <archdesc><dsc><c0x><physdesc><extent> | 0 | 0.0% |
| ***Control Access Terms*** | | 4 | 2.9% |
| **Subject Terms** | <archdesc><controlaccess><subject> | 4 | 2.9% |
| **Geographical Names** | <archdesc><controlaccess><geogname> | 4 | 2.9% |
| **Personal Names** | <archdesc><controlaccess><persname> | 0 | 0.0% |
| ***Other*** | | 2 | 1.4% |
| **[Search Term Not Found]**** | **N/A** | 2 | 1.4% |

**Note that the "Tag Label" column represents the sections as labelled in the patron view of the finding aid. The tags in the Collection Inventory were not specifically labelled, so they are represented and italicized here by their tag name.*
***Search term as applied was not actually found in the record. These were exclusively number search terms such as "25.5/8" (which is a collection number), where the system searched for each number and punctuation separately.*

*1.4 In which tags were search terms ONLY found?*

In looking at the distribution of search terms across the sections of the finding aid, one important metric was how often and where search terms were found exclusively. This metric tells the research team how often a finding aid would not have appeared to the user if the tag or section of the finding aid had been excluded. As demonstrated in *Table 5*, the Collection Inventory was the sole section in the finding aid where search terms could be found where there was only one tag in which they occurred. They represented 107 of the 140 URLs in which the search parameters were captured. This means that 76% of these URLs would not have been displayed to patrons had collection inventory not been present.

In question 3 below, the research team will break down the difference in file-level and item-level descriptions for this 76%. When looking further into the tags within the collection inventory where search terms occurred, the <scopecontent> was responsible for the vast majority of all search terms that occurred in just one tag, representing 72% of the URLs where search parameters were known. The <scopecontent> was not available in all collections – primarily occurring in just the Utah State University Football Programs and Utah State University Basketball programs. The <scopecontent> tag included the roster of players for each program. This finding aligns with the findings in research question 1.1 which found that over three-quarters of the search terms used were names. In only 6 URLs was the search term found solely in the <unittitle>. This

showed some significance compared with all other tags, excluding the <dsc><c0x><did><scopecontent>. However, this may also indicate that patron interest likely matches the content of the material more than the titles or information recorded by the processor.

Table 6. Tags where search terms are found, when only tag contains search terms

| Tag Label* | Tag Hierarchy | Number of URLs | Percentage of URLs |
|---|---|---|---|
| ***Frontmatter*** | | 0 | 0% |
| **Access Restrictions** | <archdesc><accessrestrict> | 0 | 0% |
| **Acquisitions Information** | <archdesc><acqinfo> | 0 | 0% |
| **Arrangement** | <archdesc><arrangement> | 0 | 0% |
| **Collection Number** | <archdesc><did><unitid> | 0 | 0% |
| **Content Description** | <archdesc><scopecontent> | 0 | 0% |
| **Creator** | <archdesc><did><origination><persname> | 0 | 0% |
| **Dates** | <archdesc><did><unitdate> | 0 | 0% |
| **Historical Note** | <archdesc><bioghist> | 0 | 0% |
| **Languages** | <archdesc><did><langmaterial> | 0 | 0% |
| **Preferred Citation** | <archdesc><prefercite> | 0 | 0% |
| **Processing Note** | <archdesc><processinfo> | 0 | 0% |
| **Quantity** | <archdesc><did><physdesc><extent> | 0 | 0% |
| **Related Material** | <archdesc><relatedmaterial> | 0 | 0% |
| **Repository** | <archdesc><did><repository><corpname> | 0 | 0% |
| **Restrictions on Use** | <archdesc><userestrict> | 0 | 0% |
| **Summary** | <archdesc><did><abstract> | 0 | 0% |
| **Title** | <archdesc><did><unittitle> | 0 | 0% |
| ***Collection Inventory (labelled "Detailed Description of Collection")*** | | 107 | 76% |
| *Scope and Content* | <archdesc><dsc><c0x><did><scopecontent> | 101 | 72% |
| *Unit Title* | <archdesc><dsc><c0x><did><unittitle> | 6 | 4% |
| *Unit Date* | <archdesc><dsc><c0x><did><unitdate> | 0 | 0% |
| *Container* | <archdesc><dsc><c0x><did><container> | 0 | 0% |
| *Unit ID* | <archdesc><dsc><c0x><did><unitid> | 0 | 0% |

| | | | |
|---|---|---|---|
| *Extent* | <archdesc><dsc><c0x><physdesc><extent> | 0 | 0% |
| ***Control Access Terms*** | | 0 | 0% |
| **Subject Terms** | <archdesc><controlaccess><subject> | 0 | 0% |
| **Geographical Names** | <archdesc><controlaccess><geogname> | 0 | 0% |
| **Personal Names** | <archdesc><controlaccess><persname> | 0 | 0% |
| | **Sub-total** | **107** | |
| | **Total URLs with search terms** | **140** | |

**Note that the "Tag Label" column represents the sections as labelled in the patron view of the finding aid. The tags in the Collection Inventory were not specifically labelled, so they are represented and italicized here by their tag name.*

*1.5 Did the search term category impact where it was found in the finding aid?*

As noted in *Table 3*, Last Name and First Name were the most common types of search terms used by patrons. When analyzing where search query types are most commonly found in the finding aid*, Table 7* demonstrates that Last Names and First Names were exclusively found in the Collection Inventory or <dsc>. Subject was the next most frequently occurring search query type. It was found in all three sections and in 12 tags in the finding aid: Acquisition Information, Content Description, Creator, Historical Note, Preferred Citation, Related Materials, Restrictions on Use, Summary, and Title, Scope and Content, Unit Title, and Subject Terms. “Football” and “basketball” were the most commonly used subject search terms, respectively occurring in 50% and 33% of the URLs containing subject terms.

The Collection Name, Institutional Name, Subject, and Place terms were the most likely query types to occur in multiple sections and tags of the finding aid. This is further evidenced by the fact the Collection Name often included the Subject terms most searched for (such as “football” and “basketball”) and were likely to be repeated in fields that described the history of the collection as well as how to cite and use the collection.

Table 7. Where search terms are found by search term category

| Tag Label* | Tag and Tag Hierarchy | Campuslocation | CollectionName | CollectionNumber | FirstName | Format | InstitutionalName | LastName | MiddleName | OrganizationName | Place | Subject | Title(Person) | Year | TotalNumberofSubjectTe | NumberofSearchTermCat | NumberofURLsin |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | **rms Found in Tag/Section** | **egories Found in Tag/Section** | **Which Search Term Is Present** |
| ***Frontmatter*** | | | **5** | | | | **9** | | | | **4** | **18** | **2** | | **38** | **10** | **28** |
| **Access Restrictions** | <archdesc><accessrestrict> | | | | | | | | | | | | | | **0** | **0** | **0** |
| **Acquisitions Information** | <archdesc><acqinfo> | | 3 | | | | 1 | | | | 1 | 11 | | | **16** | **7** | **14** |
| **Arrangement** | <archdesc><arrangement> | | | | | | | | | | | | | | **0** | **0** | **0** |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Collection Number** | <archdesc><did><unitid> | | | | | | 3 | | | | | | | | **3** | **3** | **3** |
| **Content Description** | <archdesc><scopecontent> | | 2 | | | | | | | | | 1 | | | **3** | **2** | **3** |
| **Creator** | <archdesc><did><origination><persname> | | | | | | 5 | | | | | 1 | | | **6** | **4** | **6** |
| **Dates** | <archdesc><did><unitdate> | | | | | | | | | | | | | | **0** | **0** | **0** |
| **Historical Note** | <archdesc><bioghist> | | 5 | | | | 6 | | | | 4 | 10 | 1 | | **26** | **10** | **20** |
| **Languages** | <archdesc><did><langmaterial> | | | | | | | | | | | | | | **0** | **0** | **0** |
| **Preferred Citation** | <archdesc><prefercite> | | 5 | | | | 9 | | | | 1 | 17 | | | **32** | **8** | **23** |
| **Processing** | <archdesc><processinfo> | | | | | | | | | | | | | | **0** | **0** | **0** |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Note** | | | | | | | | | | | | | | | | | |
| **Quantity** | <archdesc><did><physdesc><extent> | | | | | | | | | | | | | | **0** | **0** | **0** |
| **RelatedMaterials** | <archdesc><relatedmaterials> | | 3 | | | | | | | | 1 | 1 | | | **5** | **3** | **3** |
| **Repository** | <archdesc><repository><corpname> | | | | | | 8 | | | | | | | | **8** | **1** | **9** |
| **RestrictionsonUse** | <archdesc><userestrict> | | 5 | | | | 6 | | | | 1 | 17 | 2 | | **31** | **4** | **24** |
| **Summary** | <archdesc><did><abstract> | | 3 | | | | 6 | | | | 2 | 16 | | | **27** | **4** | **20** |
| **Title** | <archdesc><did><unittitle> | | 5 | | | | 6 | | | | 1 | 17 | | | **29** | **3** | **22** |
| ***Collection Inventory (labelled "Detailed Description of Collection")*** | | 1 | 2 | | 80 | 1 | 10 | 109 | 8 | 5 | 3 | 17 | 2 | 3 | **241** | **11** | **133** |
| *Scope and Content* | <archdesc><dsc><c0x><did><scopecontent> | 1 | | | 80 | | 6 | 108 | 8 | | | 8 | 2 | 1 | **214** | **8** | **114** |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Unit Title* | <archdesc><dsc><c0x><did><unittitle> | | 2 | | | 1 | 10 | 1 | | 5 | 3 | 17 | 1 | | **40** | **10** | **31** |
| *Unit Date* | <archdesc><dsc><c0x><did><unitdate> | | | | | | 3 | | | | | | | 3 | **6** | **3** | **3** |
| *Container* | <archdesc><dsc><c0x><did><container> | | | | | | | | | | | | | | **0** | **0** | **0** |
| *Unit ID* | <archdesc><dsc><c0x><did><unitid> | | | | | | | | | | | | | | **0** | **0** | **0** |
| *Extent* | <archdesc><dsc><c0x><physdesc><extent> | | | | | | | | | | | | | | **0** | **0** | **0** |
| ***Control Access Terms*** | | | 1 | | | | | | | | 4 | 1 | | | **6** | **4** | **4** |
| **SubjectTerms** | <archdesc><controlaccess><subject> | | 1 | | | | | | | | 4 | 1 | | | **6** | **4** | **4** |
| **GeographicalNames** | <archdesc><controlaccess><geogname> | | 1 | | | | | | | | 4 | | | | **5** | **4** | **4** |
| **PersonalNames** | <archdesc><controlaccess><persname> | | | | | | | | | | | | | | **0** | **0** | **0** |
| ***Other*** | | | | 2 | | | | | | | | | | | **2** | **1** | **2** |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **[Search Term Not Found]** ** | N/A | | | 2 | | | | | | | | | | | **2** | **1** | **2** |
| **Number of URLs in Which Search Term is Present** | | **1** | **5** | **2** | **80** | **1** | **12** | **109** | **8** | **5** | **4** | **20** | **2** | **3** | | | |
| **Number of Sections in Which Search Term is Present** | | **1** | **3** | **0** | **1** | **1** | **2** | **1** | **1** | **1** | **3** | **3** | **2** | **1** | | | |
| **Number of Tags in Which Search Term is Present** | | **1** | **11** | **0** | **1** | **1** | **12** | **2** | **1** | **1** | **10** | **12** | **4** | **2** | | | |

**Note that the "Tag Label" column represents the sections as labelled in the patron view of the finding aid. The tags in the Collection Inventory were not specifically labelled, so they are represented and italicized here by their tag name.*
***Search term as applied was not actually found in the record. These were exclusively number search terms such as "25.5/8" (which is a collection number), where the system searched for each number and punctuation separately.*

***Question 2: Is there a measurable difference in discoverability between finding aids described to the item-level and finding aids described to a box or folder level?***

The majority of the pageviews that were recorded came from outside the Archives West repository, and so did not have search term information in the URL that the research team could analyze. To that end, the research team also looked at simple pageviews– and how those pageviews differed between collections described to the item-level versus collections described to the box level - to determine if more detailed Collection Inventory sections resulted in greater discoverability.

*2.1 How often were the finding aids accessed?*

Given that data points in Research Question 1 indicated that the Collection Inventory plays a significant role in discovery, the research team wanted to determine if there was a difference in frequency of access between collections described at the box or folder level ("file-level") and the item-level within the Collection Inventory. As noted in *Table 8*, all 10 finding aids were accessed a total of 1,183 times during the 13-18 months they were posted online. Apart from the Adams Elementary Valentine's Tea Fieldwork collection, which was never accessed in either the file-level or the item-level finding aids, the remainder of the collections were accessed between 21 times and 860 times. The most frequently accessed collections were the Utah State University Football Programs and the Utah State University Men's Basketball Programs, which combined represented 96.8% of all pageviews.

When breaking down the pageviews by level of description, the item-level finding aids show a commanding lead for pageviews over the guides described at the file-level. As *Table 8* demonstrates, overall, 98.39% of all pageviews occurred on the item-level finding aids as compared to their file-level counterparts, showing that finding aids with an item-level description were on average 61x (or 6,100%) more discoverable than file-level description finding aids. At the high end, the item-level description for the Utah State University Football Programs collection was 106.5x (or 10,650%) more discoverable than its file-level counterpart. At the low end, the Bear River Heritage Barn Survey item-level finding aid was 3.6x (or 360%) more discoverable than the file-level finding aid. The data points show a range of 3.6x -106.5x (or 360% - 10,650%) more discoverability for item-level finding aids, with an average of 61x (or 6,100%) more discoverability, showing clearly that this level of description was consistently more successful at driving pageviews.

Table 8. Pageviews by Level of Description

| | **Pageviews by Level of Description** | | | | | |
|---|---|---|---|---|---|---|
| | **File** | | **Item** | | **Both File and Item** | |
| **Collection Name** | Pageviews | Percentage of File-level Pageviews | Pageviews | Percentage of Item-level Pageviews | **Total Pageviews** | **Percentage of Total** |
| Utah State University College Journal Index | 2 | 9.52% | 19 | 90.48% | **21** | **1.78%** |
| Utah State University Football Programs | 8 | 0.93% | 852 | 99.07% | **860** | **72.70%** |
| Adams Elementar | 0 | 0.00% | 0 | 0.00% | **0** | **0.00%** |

| | | | | | | |
|---|---|---|---|---|---|---|
| y Valentine's Tea Fieldwork | | | | | | |
| Bear River Heritage Barn Survey | 5 | 21.74% | 18 | 78.26% | **23** | **1.94%** |
| Utah State University Men's Basketball Programs | 4 | 1.43% | 275 | 98.57% | **279** | **23.58%** |
| **Total** | **19** | **1.61%** | **1164** | **98.39%** | **1183** | **100%** |

*2.2 What percentage of the days online were each finding aid accessed?*

The research team looked at the number of days that a finding aid was accessed over the course of its time posted online. Note that a finding aid could be accessed more than one time in a single day, so the number of days and the number of pageviews noted in *Table 8* above are not the same. The number-of-days metric was decided on as a way of visualizing the frequency of engagement with a finding aid. Similar to what was noted above, the item-level finding aids for the Utah State University Football Programs and Utah State University Men's Basketball Programs collections were accessed the most frequently, with the football programs accessed at least once per day on 72.38% of the days it was posted online, and the basketball programs accessed at least once per day on 42.75% of the days it was available online. (See *Table 9*.) The access for file-level descriptions for both of these collections never broke 2% of the days they were posted online. Taking the name heavy sports program collections out of the mix, the remaining collections accessed still showed a preference for the item-level description. The item-level finding aids for the Utah State University College Journal Index and the Bear River Heritage Barn Survey were accessed 3.58% and 3.23% of the days they were posted online with the file-level finding aids accessed 0.30% and 0.76% of the days online, respectively. On average, the item-level finding aids were 37.5x (or 3,750%) more likely to be accessed than the file-level finding aids.

Table 9. Number of days accessed while online

| | **Days Online** | | **Number of Days Accessed** | | **Percentage of Days Accessed** | |
|---|---|---|---|---|---|---|
| **Collection Name** | **File** | **Item** | **File** | **Item** | **File** | **Item** |
| Utah State University College Journal Index | 671 | 671 | 2 | 24 | 0.30% | 3.58% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Utah State University Football Programs | 572 | 572 | 7 | 414 | 1.22% | 72.38% |
| Adams Elementary Valentine's Tea Fieldwork | 547 | 547 | 0 | 0 | 0.00% | 0.00% |
| Bear River Heritage Barn Survey | 526 | 526 | 4 | 17 | 0.76% | 3.23% |
| Utah State University Men's Basketball Programs | 407 | 407 | 4 | 174 | 0.98% | 42.75% |
| **Total** | | | **17** | **629** | | |
| **Average** | | | | | **0.65%** | **24.39%** |

*2.3 How much time did users spend looking at each finding aid?*

On average, the item-level finding aids saw an average of 4.2x (or 420%) more time spent on the page than file-level descriptions. In a shift from previous metrics, the Utah State University College Journal Index and Bear River Heritage Barn Survey saw the most time-on-page engagement, with the item-level finding aids seeing 3.35 and 4.10 minutes on average. In another interesting comparison, the Utah State University Football Programs saw almost the same level of time engagement on file-level and item-level finding aids. (See *Table 10*.)

Table 10. Average time on page

| | **Average Time on Page (Seconds)** | | **Average Time on Page (Minutes)** | |
|---|---|---|---|---|
| | **Level of Description** | | | |
| **Collection Name** | **File** | **Item** | **File** | **Item** |
| Utah State University College Journal Index | 4 | 201 | 0.06 | 3.35 |
| Utah State University Football Programs | 60 | 58 | 1.00 | 0.97 |
| Adams Elementary Valentine's Tea Fieldwork | 0 | 0 | 0.00 | 0.00 |
| Bear River Heritage Barn Survey | 62 | 246 | 1.03 | 4.10 |
| Utah State University Men's Basketball Programs | 6 | 54 | 0.10 | 0.91 |
| **Average** | **26.27** | **111.89** | **0.44** | **1.86** |

***Question 3: How do search parameters impact discoverability for file-level description vs. item-level description?***

*3.1 How do pageviews differ for URLs with and without search parameters for file- and item-level finding aids?*

Regardless of whether search terms were present in the URL, the item-level finding aids were more likely to be viewed across the board. When search terms were present in the

URL (indicating the URL was generated by a search within the Archives West site), the item-level finding aids ranged between 4 and 13.5x (or 400% and 1,350%) more visible to users and averaged 10.2x (or 1,020%) more visibility. When search terms were not present in the URL (most likely indicating a browser search that resulted in a pageview), item-level finding aids ranged between 3 and 744x (or 300% and 74,400%) more visible to browsers than file-level finding aids, with an average of 204.2x (20,420%) more visibility. To further illustrate this point, neither the Utah State University College Journal Index nor the Utah State University Football Program collections would have been accessed at all without the item-level description, indicating that browsers were not matching the file-level finding aids for these collections with user search terms. This is substantial given the considerable number of pageviews experienced by the item-level Utah State University Football Programs finding aid. Overall, item-level finding aids were 61x (6,100%) more visible to patrons than file-level finding aids. (*See Table 11*)

Table 11. Page views for URLs with and without search parameters by collection and level of description

| URL Type | Bear River Heritage | | College Journal | | Utah State University football programs | | Utah State University men's basketball programs | | All Collections | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | File | Item | File | Item | File | Item | File | Item | File | Item | |
| No search parameters | 3 | 10 | 0 | 10 | 0 | 744 | 2 | 257 | **5** | **1021** | **1026** |
| Search Parameters Included | 2 | 8 | 2 | 9 | 8 | 108 | 2 | 18 | **14** | **143** | **157** |
| **Total** | **5** | **18** | **2** | **19** | **8** | **852** | **4** | **275** | **19** | **1164** | **1183** |

*3.2 Is there a difference in where search terms are found in file-level description vs. item-level description?*

As noted in *Table 7*, some search query types correspond with terms that occur in multiple sections in a finding aid. These include Collection Name, Institutional Name, Subject, and Place. They tend to be found in sections such as Restrictions on Use, Preferred Citation, Title, Historical Note, Summary, and Acquisition Information. All of these sections are located in the Frontmatter portion of the finding aid and during this experiment, were identical between both the item-level and file-level finding aids. Therefore, as shown in *Table 12*, it is unsurprising that the number of times search terms triggered views of the finding aids were relatively similar in the file- and item-level description finding aids. Of the URLs with search terms, only 20% matched terms in the Frontmatter, with 11.5% matching Frontmatter in the item-level finding aids and 8.5% matching Frontmatter in the file-level finding aids. The Collection Inventory or

<dsc> portion of the finding aid, though, shows that the item-level descriptions were significantly more likely to be accessed than file-level description. URLs with search terms matched the Collection Inventory 95% of the time, with 90% matching the Collection Inventory in item-level finding aids and only 5% matching the Collection Inventory in file-level finding aids.

Table 12. Section of finding aid where search terms are found in file- and item-level description finding aids

| Tag Label* | Tag and Tag Hierarchy | Item | File | Total |
|---|---|---|---|---|
| | | Number or URLs | Number of URLs | |
| ***Frontmatter*** | | ***16*** | ***12*** | ***28*** |
| **Access Restrictions** | <archdesc><accessrestrict> | 0 | 0 | 0 |
| **Acquisitions Information** | <archdesc><acqinfo> | 6 | 8 | 14 |
| **Arrangement** | <archdesc><arrangement> | 0 | 0 | 0 |
| **Collection Number** | <archdesc><did><unitid> | 1 | 2 | 3 |
| **Content Description** | <archdesc><scopecontent> | 1 | 2 | 3 |
| **Creator** | <archdesc><did><origination><persname> | 3 | 3 | 6 |
| **Dates** | <archdesc><did><unitdate> | 0 | 0 | 0 |
| **Historical Note** | <archdesc><bioghist> | 11 | 9 | 20 |
| **Languages** | <archdesc><did><langmaterial> | 0 | 0 | 0 |
| **Preferred Citation** | <archdesc><prefercite> | 11 | 12 | 23 |
| **Processing Note** | <archdesc><processinfo> | 0 | 0 | 0 |
| **Quantity** | <archdesc><did><physdesc><extent> | 0 | 0 | 0 |
| **Related Materials** | <archdesc><relatedmaterials> | 1 | 2 | 3 |
| **Repository** | <archdesc><repository><corpname> | 4 | 5 | 9 |
| **Restrictions on Use** | <archdesc><userestrict> | 13 | 11 | 24 |
| **Summary** | <archdesc><did><abstract> | 11 | 9 | 20 |
| **Title** | <archdesc><did><unittitle> | 11 | 11 | 22 |
| ***Collection Inventory (labelled "Detailed Description of Collection")*** | | ***126*** | ***7*** | ***133*** |
| *Scope and Content* | <archdesc><dsc><c0x><did><scopecontent> | 114 | 0 | 114 |
| *Unit Title* | <archdesc><dsc><c0x><did><unittitle> | 24 | 7 | 31 |

| | | | | |
|---|---|---|---|---|
| *Unit Date* | <archdesc><dsc><c0x><did><unitdate> | 1 | 2 | 3 |
| *Container* | <archdesc><dsc><c0x><did><container> | 0 | 0 | 0 |
| *Unit ID* | <archdesc><dsc><c0x><did><unitid> | 0 | 0 | 0 |
| *Extent* | <archdesc><dsc><c0x><physdesc><extent> | 0 | 0 | 0 |
| ***Control Access Terms*** | | ***3*** | ***1*** | ***4*** |
| **Subject Terms** | <archdesc><controlaccess><subject> | 3 | 1 | 4 |
| **Geographical Names** | <archdesc><controlaccess><geogname> | 3 | 1 | 4 |
| **Personal Names** | <archdesc><controlaccess><persname> | 0 | 0 | 0 |
| ***Other*** | | ***2*** | ***0*** | ***2*** |
| **[Search Term Not Found]**** | N/A | 2 | 0 | 2 |
| | **URLs Accessed** | **128** | **12** | **140** |

**Note that the "Tag Label" column represents the sections as labelled in the patron view of the finding aid. The tags in the Collection Inventory were not specifically labelled, so they are represented and italicized here by their tag name.*
***Search term as applied was not actually found in the record. These were exclusively number search terms such as "25.5/8" (which is a collection number), where the system searched for each number and punctuation separately.*

**Discussion**

The major themes from this research and analysis include: 1) most search terms used were names, 2) the Collection Inventory or <dsc> portion of finding aids drove pageviews, and 3) the lack of content within the Collection Inventory section of a finding aid negatively impacted discoverability. These observations suggest that minimal or lacking description in the Collection Inventory of a finding aid adversely effects discoverability which, in turns, impacts user access to collections.

When answering the question "**What type of search terms are most commonly used by patrons and where are they found in the finding aid?"** the research team found that search terms were predominantly personal names representing as much as 77.86% of all URLs with search parameters. Subject terms were only found in 14.29% of URLs with search parameters. This suggests that in order to match prominent user search patterns, archival descriptive practices can benefit from incorporating personal names wherever feasible. The research team also found that search terms were overwhelmingly found in the Collection Inventory, with 95% of all search terms found in this section. Adding to this picture, the <scopecontent> and <unittitle> tags in the Collection Inventory were also the only tags in which search terms were ONLY found, meaning that if the Collection Inventory had not been present, the finding aid would not have been visible to the user. This occurred for 76% of the URLs where search parameters were found. For all other tags in which search terms were found, those same terms were

also located in other tags. These findings suggest that the content found in the Collection Inventory was often unique within the finding aid and that it more closely matched the terms that patrons were using when searching for content. Looking even further into the content of the Collection Inventory found that the addition of personal names in the <scopecontent> tag made it the most likely tag to match search terms. More emphasis on Collection Inventory or the inclusion of personal names wherever appropriate could improve the discovery of finding aids.

When examining the findings for the question "**Is there a measurable difference in discoverability between finding aids described to the item-level and finding aids described to a box or folder level?**" the research team found that finding aids with more robust Collection Inventory descriptions were more discoverable to patrons. Finding aids with item-level descriptions were, on average, 61x (or 6,100%) more discoverable than their file-level counterparts, with item-level descriptions ranging from 3.6-106.5x (360% - 10,650%) more discoverable, depending on the collection. This was further reinforced by the fact that 98.39% of all pageviews across this entire research project occurred in finding aids with the item-level description in the Collection Inventory. Item-level finding aids were also, on average, accessed on 37.5x (or 3,750%) more days and users spent 4.2x (420%) more minutes on the finding aid page than finding aids with file-level descriptions. These findings suggest that beyond the simple presence of a Collection Inventory, the extent to which the material is described significantly impacts discoverability.

Finally, when the researchers looked further at the question "**How do search parameters impact discoverability for file-level description vs. item-level description?"** to determine how the first two research questions intersected, they found that item-level description was the single biggest driver of pageviews. When search parameters were known, item-level finding aids averaged 10.2x (or 1,020%) more visibility for patrons. More significantly, though, for all pageviews where the search parameters were unknown (and therefore the pageviews were driven primarily by browser traffic), item-level finding aids averaged 204.2x (20,420%) more visibility to patrons. Two of the collections in this research project, including the Utah State University Football Programs, the most highly accessed collection, would not have been visible to browser traffic ***at all*** if they did not have an item-level description. In the first research question, the research team identified that the Collection Inventory included the search term for 95% of URLs with search parameters. When breaking that down by item-level vs. file-level descriptions, the Frontmatter in both types of finding aids showed relatively the same frequency of access. The item-level finding aids, though, showed a substantial increase in access in the Collection Inventory section, with 90% of URLs matching the item-level finding aids and 5% matching the file-level finding aids. This was particularly found in the <scopecontent> tag, but also in the <unittitle> tag.

The data in this research project showed that personal names are a driving category of search terms for archival materials. The Utah State University Football Programs and Utah State University Basketball Programs were both name-heavy collections and resulted in increased discovery for patrons looking for individuals. Interestingly, the personal names searched for by patrons were usually unique (not repeated across

searches) and did not include the more prominent or well-known sports names available in the programs. This showed a wide range in the interest for personal names. These collections also showed regular and consistent patron engagement over the course of the days they were online, indicating that these searches were not one-time or isolated events. Additionally, the Adam's Elementary Valentine's Tea Fieldwork and the Bear River Heritage Barn Survey collections both included personal names, albeit at a much-reduced rate. However, they did not garner anywhere near the same interest, which presents an interesting question about whether all personal names drive discovery or merely specific categories of personal names. This is an avenue the research team will be exploring in the future.

The research team also noted that personal names were not often found in the Control Access section because 4 of the 6 collections did not have this content due to policy decisions on minimal processing at the time they were created. Descriptive practices that incorporate more authorized name headings may be a good step for increasing not just the discoverability of finding aids, but their interoperability with finding aids from other institutions. However, some collections, such as the sports programs, may provide too many personal names to feasibly develop authorized headings for each and every name present. In cases like this, simply listing the names in the Collection Inventory may at least provide a measure of discoverability for patrons, even if does not make the finding aid interoperable with other institutions.

In looking at the daunting task of doing item-level description and including personal names wherever present, it is unlikely that many institutions, including Utah State University Libraries, will have the time or staffing to develop detailed collection descriptions. USU Libraries will lean on a previously developed process, described in Woolcott et al.[17] that enhances minimally described finding aids with more robust metadata created during the digitization of items. However, it should be noted that this is not an ideal workflow because the decision to digitize collections may be dependent on the usage of the collection which may, in turn, be highly dependent on the discoverability of the items in that collection.

## Conclusion

The data points from this research project indicate that the unique information contained in the Collection Inventory or <dsc> section in a finding aid plays a crucial role in the discovery process. More detailed Collection Inventories, developed to the item-level, resulted in substantially higher discovery and user engagement than Collection Inventories developed to the file-level. On average, finding aids that were identical in every regard except the Collection Inventory, saw 61x (or 6,100 %) more traffic to the item-level description version over the file-level description. Even excluding unusually detailed finding aids, the use of item-level description resulted in a minimum of 3.6x (or 360%) increase in pageviews. In some instances, the finding aids in this project were invisible to browser searches as well as search queries within the repository without item-level descriptions. Additionally, the majority of user traffic was driven towards name-heavy collections. Name queries represented as much as 77.86% of all searches where the search terms could be identified. In this research project, personal names that matched user queries were exclusively found in the Collection Inventory and did not surface in the other sections of the finding aid.

**Acknowledgements**

The authors would like to acknowledge the substantial contributions of the catalogers, cataloging assistants, and student technicians of the Cataloging and Metadata Services unit of Utah State University Libraries. Their efforts were significant in contributing to the description and creation of the finding aids for this project. In particular, the authors would like to recognize Sara Skindelien, who encoded all the finding aids in this project, as well as Bryn Larsen, Maddison Gardner, and Melanie Shaw who transcribed the names of the players for the Utah State University Men's Basketball Program collection and the Utah State University Football Programs collection. The authors would also like to acknowledge Randy Williams (former Fife Folklore Archivist) and Bob Parsons (former University Archivist) for identifying the collections for this project and their thoughtful contributions throughout the project. And finally, the authors would like to thank Terri Jordan (current Fife Folklore Archivist) and Kelly Rovegno (current University Archivist) who came on board during a pandemic and continued the project, providing vital input and feedback on the final analysis.

**References**

---

[1] Woolcott, L., Payant, A., & Skindelien, S. (2016). Partnering for Discoverability: Knitting Archival Finding Aids to Digitized Material Using a Low-Tech Digital Content Linking Process. *Code {4} lib*, (34). https://journal.code4lib.org/articles/11997

[2] 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

[3] Daniels, M., & Yakel, E. (2010). Seek and You May Find: Successful Search in Online Finding Aid Systems. *The American Archivist*, *73*(2), *535-568*. https://doi.org/10.17723/aarc.73.2.p578900680650357

[4] Yang, L. (2016). Metadata effectiveness in internet discovery: An analysis of digital collection metadata elements and internet search engine keywords. *College & Research Libraries. 77(1):7-19.* https://doi.org/10.5860/crl.77.1.7

[5] Bronn, M., Proffitt, M. & Washburn, B. (2013). Thresholds for discovery: EAD tag analysis in ArchiveGrid, and implications for discovery systems. *Code4lib, Journal 22.*

[6] Washburn, B. (2021, August 9). *How well does EAD Tag Usage Support Finding Aid Discovery?* Hanging Together. Retrieved November 16, 2021, from https://hangingtogether.org/how-well-does-ead-tag-usage-support-finding-aid-discovery/.

[7] 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

[8] Meissner, D., & Greene, M. A. (2010). More Application while Less Appreciation: The Adopters and Antagonists of MPLP. *Journal of Archival Organization*, *8*(3–4), 174–226. https://doi.org/10.1080/15332748.2010.554069

[9] Hauck, J., Krause, R.S., & Herzinger, K. (2019). MPLP Ten Years Later: The Adventure of Being among the First. *SPU Works, 136.*

[10] Edmunson-Morton, T. (2007) SAA 2007 Session #307. *Archival Musings* (blog), Retrieved November 16, 2021, from http://temarchivalmusings.blogspot.com/2007/09/saa-2007-session-307.html

[11] Crowe, S., & Spilman, K. (2010). MPLP @ 5: More Access, Less Backlog?. *Journal of Archival Organization, 8*(2), 110-113.

[12] Parilla, L.L., Morgan, R. and Fidler, C. (2017). Excavating Archival Description: From Collection to Data Level. *Digital Library Perspectives*, *33(3),* 195-202. https://doi.org/10.1108/DLP-11-2016-0043

[13] Aleman, S., Christner, C., & McGee, M. (2017). *Why Can't I Find You? A User-Centric Study and Recommendations for Location Archival Information.* Retrieved November 19, 2021, from https://www.mollykmcgee.com/images/docs/usercentric_study_archival_finding_AlemanChristnerMcGee.pdf

[14] Higgins, S., Hilton, C., & Dafis, L. (2014). Archives context and discovery: Rethinking arrangement and description for the digital age. Paper presented at *2nd Annual Conference of the International Council on Archives, Girona, Spain.*

[15] Zhang, J., & Mauney, D. (2013). When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation. *The American Archivist*, *76*(1), 174–195. https://doi.org/10.17723/aarc.76.1.121u85342062w155

[16] Therrell, G. (2018). More Product, More Process: Metadata in Digital Image Collections. *Digital Library Perspectives, 35(1), 2-14*. https://doi.org/10.1108/DLP-06-2018-0018

[17] Woolcott et al. (2016)