

Ayla Toussaint. Finding a Way Through the Crowd: How Keyword Choices Affect Discoverability in Crowdsourced Archival Tagging. A Master's paper for the M.S. in L.S. degree. November, 2018. 60 pages. Advisor: Ryan Shaw

This study explores the challenges archivists face when determining how to structure a crowdsourced tagging initiative in their collections. Specifically, the study aims to research the differences in controlled vocabulary tagging and folksonomy tagging in volunteer based digital archival tagging initiatives. The current literature focuses on the merits of crowdsourced metadata in its various forms, but few sources explore which formats lead to better discoverability.

In order to begin to answer this question, five library, museum, and archives professionals were interviewed and asked to discuss their institution's crowdsourcing projects in depth. Their answers were then mined for overarching themes and insights into crowdsourcing and vocabulary type. In the end, it was discovered that there is no one correct vocabulary system for crowdsourcing, but that by answering key questions about specific institutions, collections, and volunteers, a unique approach can be created for each new project to ensure the best outcome.

Headings:

Archives

Crowdsourcing

Folksonomies

Subject headings – Use studies

Tags (Metadata)

FINDING A WAY THROUGH THE CROWD:
HOW KEYWORD CHOICES AFFECT DISCOVERABILITY IN CROWDSOURCED ARCHIVAL
TAGGING

By Ayla Toussaint

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

November, 2018

Approved by:

Ryan Shaw

Table of Contents

Table of Contents	1
Introduction	2
Literature Review.....	8
Methodology.....	20
Findings.....	24
Discussion	40
Conclusions.....	46
Appendix A: Email Form.....	48
Appendix B: Interview Question Template	49

Introduction

Processing collections as they come into archives is becoming more time intensive and unmanageable as collections continue to grow in size and complexity. This issue has been exacerbated by the rise of born digital materials and the push for digital access to analog materials. The backlog of unprocessed materials we are beginning to see in archives is resulting in “lost” or forgotten items of historical significance. When these materials are finally “found” by researchers, they can tremendously change how we understand past events and historical figures. Even processed collections can house “lost” items to be “rediscovered” by a lucky historian, as the vast majority of collections are not processed at the item level. Although the increased dependency on machines and digital discovery has posed a problem for archivists, it could also hold the answer to processing such large collections in a way that aids discovery and use. Crowdsourcing collections for transcription and tagging has become increasingly prevalent in the past few years, with volunteer numbers increasing every day.

This paper aims to explore the increase of crowdsourced tagging in archival collections, specifically in digitized collections available online. Specifically, how crowdsourced user-generated/folksonomy tagging on item levels affect discoverability and accessibility as compared to crowdsourced controlled vocabulary tagging on archival materials. In order to dive into this exploration, we must first look at the

background of the issues in the field and define our terms. An in-depth analysis of the literature can be seen in the "Literature Review" section of this paper. As an introduction, this section will simply set the stage with the basics and the archival background for my question. After setting the scene, I will introduce my question more in depth and discuss the limitations and scope of this study.

Background

As long as archives have existed, questions of how to store increasing amounts of materials have haunted archivists and caused them to reassess their collection policies. This trend is not new, but has seen a dramatic, exponential increase in relevance after both the advent of typewriters and mass printing, and, more recently, the rise of the digital age.¹ The most obvious question may be "where do we put everything?", but the more difficult one is "how do we process everything in order to ethically preserve its worth and make it reasonably accessible?". A favorite saying among archivists is there is no such thing as benign neglect.² This means that leaving records in storage without properly processing them is by definition harmful. While for some collections this can be physical harm, such as deterioration or other damage, for others it is more abstract. When we leave collections in storage waiting to be processed, they remain in a sort of limbo in which they cannot fulfill their purpose. Researchers cannot access the materials and their information is lost in the proverbial sea of paper.

The increase of records correlates to the growing backlog of unprocessed archival collections. Processing a collection properly takes time and as the field currently stands, the influx of materials into archives is happening faster than archivists can process them. This is a multifaceted issue, but the biggest factor on the archivists' side is their institutions' available resources. Oftentimes, archives simply do not have the

manpower, time, or money to fully describe all of their collections as they come in. This leads to a backlog of unprocessed materials and many partially processed collections.

As the creation of data continues to rise exponentially, many institutions have adopted a MPLP (more product, less processing) approach to description. This helps get collections into finding aids but does not give users much to go on in way of description.³ Frequently, collections will remain minimally processed until a research request is made, at which point the collection will be looked at more thoroughly.⁴

This is where the introduction of crowdsourcing and Web 2.0 technologies comes into play. Archives are recognizing that they can combat issues of limited resources by calling on the public as volunteers to help process collections. This is seen in transcription projects and online tagging games for digitized papers, photographs, and other materials.⁵ These projects ask volunteers to tag objects or provide transcriptions of documents in order to go up a level or earn points towards virtual prizes. The gaming feel makes the work fun and fulfilling for volunteers, while the content sparks their interest in history. Archivists compile the responses of the volunteers and check for accuracy before incorporating the data into finding aids and item descriptions. This process allows archivists to process their materials and get them out to the public much faster than if they were to endeavor to do it themselves. In addition, the majority of the labor is freely given, which means that it does not tax their funds.

This type of tagging and processing is useful, but research has found that the keywords generated by the public vs scholars vs archivists are all different.⁶ It is unclear how this affects discoverability. This inconsistency across tagging is why libraries and archives often utilize controlled vocabularies when indexing and processing materials. In a controlled vocabulary, one definitive word is chosen to represent all synonyms or

like terms. For example, Cat is used for cat, feline, cats, kitty, etc. This means that only one search is required to find all materials relating to cats, instead of searches with all related terms.⁷ This study aims to explore how controlled vocabulary might help or hinder discoverability in crowdsource tagged materials as well as how implementing a controlled vocabulary affects volunteers.

Research Question

My research question revolves around this new way of processing collections at the item level. **How does crowdsourced user-generated/folksonomy tagging on item levels affect discoverability and accessibility as compared to crowdsourced controlled vocabulary tagging on archival materials?** Although there is no doubt that crowdsourcing processes collections faster than archives could with their own limited resources, the question of how this aids in discoverability is not as clear. Users are not always experts on the collections and may miss key details or may use outdated or obscure words to tag an item, hindering its future searchability.⁸ I would like to explore how integrating controlled vocabularies or thesauruses into crowdsourcing initiatives would affect volunteer numbers and usability on all sides.

Scope/Limitations

This is quite a large topic, so I am going to briefly list branches that are out of my scope for this study. I will not be looking into how archives would receive funds to digitize their materials in preparation for this crowdsourcing. I also will not be able to note how these collections are screened for PII or sensitive material. I hope that my findings can be translated into the physical archives with in person volunteers but looking at this specifically will be out of scope as well. Interviewing volunteers and users of the systems will not be possible but could be done as a follow up to this paper.

Finally, I will not be studying the software or computing that would be required to make a digital thesaurus or key word list for volunteers to utilize, but instead the implications of such practices.

Glossary of Important Terms

- Crowdsourcing: using the general public to complete large projects on a volunteer basis that would otherwise not be able to be done due to limited resources.
- Controlled Vocabulary: a pre-fabricated list of chosen/approved descriptive words to be used on a collection.
- Tagging: giving keyword tags to specific item level files/objects.
- Folksonomy: user-generated tags or keywords created for the description of collection materials.
- Discoverability: How easy or difficult it is to find an item using traditional searching methods such as Boolean or keyword searches in online or physical repositories.
- Accessibility: How easily accessible an item is based on user experiences from general search all the way to item level research and publication of findings.
- Unprocessed: The state of archival holdings that have been acquisitioned and accepted into the repository but have not been logged or described in the archives' databases or finding aids.
- Item Level: A type of processing in which the contents of a collection have been described individually, item by item, so that there exists a complete list of every item and each item is searchable and discoverable either digitally or in paper finding aids.

¹ Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (September 1, 2005): 208–63; JoAnne Yates, *Control Through Communication: The Rise of System in American Management* (JHU Press, 1993).

² Denise Anthony, in discussion with author, Spring 2017; Camille Tyndall Watson, in discussion with author, Fall 2017.

^{3 3} Mark Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (September 1, 2005): 208-63; Terry Cook, "'Many are Called, but Few Are Chosen': Appraisal Guidelines for Sampling and Selecting Case Files," *Archivaria* 32 (Summer 1991): 25-50.

⁴ Matthew Turi, interview by Ayla Toussaint, February 7th, 2018.

⁵ Luke Barrington, Douglas Turnbull, and Gert Lanckriet, "Game-Powered Machine Learning," *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 17 (2012): 6411-16; Mary Flanagan and Peter Carini, "How Games Can Help Us Access and Understand Archival Images," *The American Archivist* 75, no. 2 (2012): 514-37; Liliana Melgar Estrada et al., "Time-Based Tags for Fiction Movies: Comparing Experts to Novices Using a Video Labeling Game," *Journal of the Association for Information Science & Technology* 68, no. 2 (February 2017): 348-64.

⁶ Marija Petek, "Comparing User-generated and Librarian-generated Metadata on Digital Images," *OCLC Systems & Services: International Digital Library Perspectives* 28, no. 2 (May 25, 2012): 101-11; J. Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept," *New Review of Hypermedia and Multimedia* 12, no. 1 (June 1, 2006): 83-105.

⁷ Robert Losee, "Chapter 3: Representation," in *Information from Processes: About the Nature of Information Creation, Use, and Representation* (Springer, 2012): 85-108.

⁸ Edward Benoit, "#MPLP: A Comparison of Domain Novice and Expert User-Generated Tags in a Minimally Processed Digital Archive" (The University of Wisconsin - Milwaukee, 2014).

Literature Review

Now that we have grounded ourselves in this study, we must take a detailed look at the surrounding literature. Despite the fact that digital crowdsourcing for archival collection metadata is a relatively new phenomenon, a plethora of studies have already been conducted and published on the efficacy of their use and design models within the fields of library science and digital humanities. In addition to these studies, we will also look into the classification schemes and protocols that archives have used up to this point, in order to ground ourselves in the literature. We must know the past in order to understand the present and imagine the future.

Let us first dive into controlled vocabularies and how they relate to other ways of describing records such as thesauri, taxonomies, folksonomies, and keywords. Controlled vocabularies are common shop talk in library science, but what do we actually mean when referring to them? According to Madely du Preez, controlled vocabularies are “restricted lists of words that are used for indexing.”⁹ Essentially, controlled vocabularies are a list of words we as information professionals use to tag sources in a uniform manner. Using a controlled, or previously devised, list of approved terms, we can be sure that all items relating to the idea of cats for example come up when the controlled vocabulary term “Cat” is searched.¹⁰ Thesaurus and controlled vocabulary are often used as interchangeable terms, but there are some

differences, as a controlled vocabulary does not necessarily have to employ a full thesaurus.

A thesaurus is a similar type of cataloging helper that takes a controlled vocabulary and connects words based on relationships to one another. Words are connected through related, broader, and narrower terms in order to guide researchers to their ideal search.¹¹ This idea of related, broader, and narrower terms is important when defining the distinction between thesauri and taxonomies. A taxonomy is similar to a thesaurus, but the relationship between the terms is hierarchical. Taxonomies form a tree like structure of terms. Madely du Preez explains the differences well.

“For example, a given term in a thesaurus may or may not have a broader/narrower term relationship with another term. Taxonomies, on the other hand, have a strict hierarchical structure. All terms in a taxonomy belong to a single large hierarchy which encompasses all concepts of a certain class, category, or aspect. Furthermore, terms in a thesaurus can have an equal relationship with other terms, as for instance with dog breeds and cat breeds. Because of the strict hierarchical structure of taxonomies, however, there can be no equal relationships in them.”¹²

Taxonomies and Thesauri are closely tied to the idea of controlled vocabularies, as they both employ the idea of a controlled list of authoritative terms to aid in searchability and discoverability of materials.

Folksonomies and keywords are on the opposite side of the spectrum from these three indexing structures, though they function in the same manner. Folksonomies gained their name through the combination of the words “folk” and “taxonomy”. They are, in effect, a type of taxonomy created by the people using their words, not those of scholars. In the digital era, folksonomies and crowdsourced tagging go hand in hand.¹³

As Madely du Preez writes, folksonomies are “the free allocation of keywords by anyone and everyone in an information system.”¹⁴ In order to fully understand this succinct definition, we must define what constitutes a keyword. A keyword functions in the same way that a subject heading does in a controlled vocabulary, except that it is not standardized across a collection. Keywords are uncontrolled tags that are assigned to records to aid in discoverability. Keywords are often utilized in databases to allow authors to tag their articles in their own vocabulary. This allows for specificity but can also limit discoverability as not all authors with similar subjects will employ the same keywords. Because of this, databases will often use both methods of tagging to allow users more options and freedom in their searches.¹⁵

As can be seen, deciding which type of description to use on a collection can be a difficult task. In recent years, many databases have been changing their protocols, either to aid in discoverability or to save resources (and often both). This has been a point of discussion at annual meetings such as ALA. In one such roundtable discussion, leaders in databases and publishing explained that metadata in the form of thesauri and controlled vocabulary are too expensive in relation to the information they describe and only help seasoned searchers find items more quickly than full text and abstract searches. Due to this, we have been seeing a steady decrease in thesauri use in databases as preference for full text and keyword searches continue to dominate.¹⁶ This is not the case everywhere; however, more and more libraries and archives have adopted social tagging and keyword searching in addition to their traditional cataloging and describing methods. This double pronged approach seems to aid in discoverability and is popular in the field as a whole.¹⁷

Before the prevalence of the internet, library catalogs and archive finding aids were how patrons found resources. With these systems, carefully crafted and applied controlled vocabularies created by information professionals were the norm in discoverability standards. As the web continued to grow, however, we began to see an increase of user generated metadata and homegrown databases for users to share and make data accessible. Sites such as Flickr, LibraryThing, and What's on the Menu? began proliferating user generated information and started the trend of Web 2.0 capabilities being used in content discovery.¹⁸ As the web continued to grow, and users became more proficient and eager to use these new technologies, libraries and archives began to incorporate them into crowdsourcing initiatives and discoverability updates. The first iteration of this trend was seen in library catalogs, with social tags and keywords appearing alongside librarian-created metadata. These initiatives were met with some success and some reluctance but are now being accepted as common-place, especially in the realm of public libraries.¹⁹ The reluctance in many of these cases stemmed from the feeling of users that keywords created by their peers would not be as reliable as those generated by trusted officials (such as librarians).²⁰ This is a common trend seen throughout the literature on crowd-sourcing, but there is no concrete answer to the reliability of crowd-sourced tags. Despite this concern, crowdsourcing in the context of social tagging has proliferated the field, with archives and museums opening their collections to social tags and user-generated metadata as well.²¹ This is not to say that all Web 2.0 focuses on social tagging. On the contrary, Web 2.0 within the library and archives realm is an enormous topic including ideas such as library outreach, user contributed collections, and other types of activities facilitated by the web and social

media. Because of the scope of this study, only a small subsection of this larger topic is being examined here.

The next iteration of Web 2.0 and social tagging took the form of crowd-sourcing through tagging game interfaces. Essentially, cultural institutions created interactive games to encourage users to tag their collections. These games can be complex and mask the tagging behind a quick and exciting task or can be as simple as giving badges to users as they complete certain tasks.²² These incentives have been shown to increase return users and help bring in new participants as well, speeding up the tagging process.²³ Unfortunately, this gaming aspect and its emphasis on speed can sometimes negatively affect the accuracy of the user tags generated. This is an issue that is commonly cited as reason to discount crowdsourcing efforts, as reliability is spotty in quick-paced games where quantity is more important than quality.²⁴

Whether through complex games or simple interfaces, archives have begun to implement crowdsourcing technologies in order to help process backlogged materials. Libraries also use collaborative tagging in their catalogs and digital collections, with varied success.²⁵ Many of the literature on these initiatives cross over institutional lines, such as Dawson's article "Crowdsourcing the library and Archive," which explores several different projects in the field presented at the conference of the same name.²⁶ Two of the most popular types of tagging in GLAMS (Galleries, Libraries, Archives, Museums) are coding photo collections with keywords and the transcription of historical documents. Galleries and Museums have been using crowdsourcing as a form of outreach and continued participant engagement for their collections. Methods of this include social tagging in exhibits. Social tagging in this context refers to allowing users to create their own tags for pieces they see during their visit. These tags can be location

based or can be applied after the visit. These options increase both active participation within an exhibit as well as continued engagement after leaving the institution. Several studies have explored these efforts, including Linda Zajac's "Social Metadata Use in Art Museums: The Case of Social Tagging" and Gunho Chae's "Linking and Clustering Artworks Using Social Tags: Revitalizing Crowd-sourced Information on Cultural Collections."²⁷ These two studies focus on art exhibits, but this technology can be applied to a variety of materials.

The Smithsonian Institution has been a leader in the field in regard to crowdsourcing and user involvement with their collections. This can be seen in a number of their initiatives such as the Smithsonian Gardens Public Tagging Initiative, where users are encouraged to add metadata to digital photo collections, and their various transcription efforts.²⁸ Transcription crowdsourcing has become a successful trend recently, with various programs across different subject matter. One recent study explored the accuracy of transcription and annotation of ancient Quranic texts by expert users. This project is fascinating as it pulls ideas from crowdsourcing technologies and combines several types of metadata in one tool.²⁹ On the opposite end of the spectrum, the University College London used volunteers to transcribe materials from their Jeremy Bentham digitized collections in their "Transcribe Bentham" project. The project met with marked success in both increasing public participation with archival collections and increasing the usability of the collections.³⁰ This is truly an international phenomenon, as the National Archives also hosts online user transcriptions, calling their users Citizen Archivists.³¹

This idea of Citizen Archivists has increased in popularity over the last several years. By asking for public support through contributions with traditionally archival

duties, we are bringing users into the archives. In these projects we give the public certain responsibilities traditionally reserved for archivists in the care and keeping of collections. It is only fair, therefore, that we give our volunteers a title to represent this level of trust. Jan Zastrow discusses the uses of crowdsourcing with Citizen Archivists in her article "The Digital Archivist. Crowdsourcing Cultural Heritage: 'Citizen Archivists' for the Future." Although there are hurdles to overcome, she believes that crowdsourcing and trusting the public can increase discoverability of archival items which would otherwise remain unprocessed.³² The idea that users can hold meaningful information and contributions for archives has been spreading, with crowdsourcing seen in the National Archives of Estonia, as described by Tarvo Kärberg and Koit Saarevet in their article "Transforming User Knowledge into Archival Knowledge."³³ This shift in ideology from users as consumers and archivists as producers/keepers of information to archives as a collaboration between archivists and citizens is a driving force behind crowdsourcing efforts.

Although we have witnessed this slow evolution in how professionals view users from simply information seekers to information experts in some areas, there is still concern over the reliability of crowdsourced metadata. Virtually every study concerned with crowdsourcing in GLAMS touches on how certain technologies and procedures affect reliability of the information. The studies mainly focus on keywords, but some also explore transcriptions and annotations. Many scholars seem to believe that the method, although effective in promoting collections, can actually be detrimental to the long-term usability of materials. This idea is explored by Janelle Varin, for example, in her article "iTunes Metadata and Classical Music: Issues and Solutions for Crowdsourced Metadata in iTunes." This piece is centered on the issue of incorrect

metadata coded when non-experts attempt to help classify music they are unfamiliar with.³⁴ Articles such as this touch on the differences between users and expert users in crowdsourcing initiatives. A study of a popular labeling game software use on film archives suggests that novice and expert tagging can have significant differences that will ultimately affect discoverability if left unchecked.³⁵ These studies acknowledge that, when done correctly, crowdsourced tags can significantly increase discoverability; however, steady reliability is too absent to justify its use. This view is seen particularly in Pamela Meyer's article "Like a Box of Chocolates: A Case Study of User-Contributed Content at Footnote."³⁶ Reliability means more than just the presence of accurate tags. In order for an item to be reliably tagged, it must be relevant to the content and context of the item and give an accurate description, but it must also be useful to researchers. Many studies have found tags to be too personal to be helpful for other users or on the opposite end, too broad to measurably aid in discoverability (i.e. "history" in a historical archive).³⁷

As touched on earlier, one of the fundamental issues archivists have in regard to crowdsourcing is with folksonomies and user-generated key words as opposed to controlled vocabulary use. This is a multifaceted issue because the field is still divided on the use of folksonomies vs. controlled vocabularies. One of the main points brought up against folksonomies is that the use of synonyms causes confusion. These scholars posit that folksonomies lack a mechanism for easy disambiguation of similar terms. Similarly, the use of colloquialisms can cause confusion within the user population, especially when the user base is diverse. On the opposite side, however, archivists point out that users may find keywords using colloquialisms to be helpful in their searches, especially if the users do not have the scholarly vocabulary which would be used

otherwise. They also note that, although tagging may differ in the beginning of a project, volunteers tend to take queues from each other when tagging and will mimic expert user behaviors.³⁸ Questions about the reliability of folksonomy tagged collections versus those tagged with controlled vocabulary is at the heart of this study.³⁹

Because there is such division on this issue, there have been several studies conducted looking into the differences between librarian/archivist coded materials and volunteer coded materials. The goal of these studies is to ascertain whether crowdsourced tags can be trusted. Most of these studies agree that there is a marked difference between the two in relation to tagging styles; however, this does not necessarily mean that discoverability is affected. In fact, several of these studies have found that, although the tags are different, the librarian vs user generated metadata had no significant effect on discoverability.⁴⁰ This should be good news for those in favor of crowdsourcing archival collections, but there are still drawbacks. Although tags in these studies did not change discoverability, users of the collections have nevertheless expressed reservations. Their concerns focus on their perceived trustworthiness of the tags. As has been seen previously, users want experts to describe collections as opposed to their peers. This is especially true in archives, where many users of the materials are scholars and heavy researchers.⁴¹

So where does this all leave us? Archives have embraced the utility of the internet for their collections and are continuing to implement new projects as trends develop. These trends often affect how collections are organized with focus on vocabulary choice and layout. Vocabulary choices and how we choose to organize our collections is unique to each institution, with merits and hurdles to each system. GLAMS have come to appreciate the effectiveness of Web 2.0 technologies in connecting

with their patrons and continue to think up new and creative ways of bringing materials to the user and vice versa. The field is still divided on the reliability of crowdsourced tagging; however, recent studies are revealing the merits of these types of projects.

⁹ Madely du Preez, "Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?," *The Indexer* 33 (January 1, 2015): 29.

¹⁰ Robert Losee, "Chapter 3: Representation," in *Information from Processes: About the Nature of Information Creation, Use, and Representation* (Springer, 2012): 85-108.

¹¹ Madely du Preez, "Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?," *The Indexer* 33 (January 1, 2015); Suzanne S. Bell, *Librarian's Guide to Online Searching*, 4th ed. (Libraries Unlimited, 2015): 33.

¹² Madely du Preez, "Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?," *The Indexer* 33 (January 1, 2015): 33.

¹³ Madely du Preez, "Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?," *The Indexer* 33 (January 1, 2015).

¹⁴ Madely du Preez, "Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?," *The Indexer* 33 (January 1, 2015): 35.

¹⁵ American Library Association, "Speaking Technically," *American Libraries* 39, no. 7 (2008): 54-57; Stephanie Brown, "INLS 501" *University of Chapel Hill School of Information and Library Sciences* (Spring 2018).

¹⁶ American Library Association, "Speaking Technically," *American Libraries* 39, no. 7 (2008): 54-57; Stephanie Brown, "INLS 501" *University of Chapel Hill School of Information and Library Sciences* (Spring 2018).

¹⁷ Sarah Hayman and Nick Lothian, "Integrating User Tagging and Controlled Vocabularies for Australian Education Networks," *IFLA Conference Proceedings*, November 2007, 1-27; Corinne Jörgensen, Besiki Stvilia, and Shuheng Wu, "Assessing the Relationships Among Tag Syntax, Semantics, and Perceived Usefulness," *Journal of the Association for Information Science & Technology* 65, no. 4 (April 2014): 836-49; Margaret E. I. Kipp and D. Grant Campbell, "Searching with Tags: Do Tags Help Users Find Things?," *Knowledge Organization* 37, no. 4 (October 2010): 239-55.

¹⁸ Margaret E. I. Kipp and D. Grant Campbell, "Searching with Tags: Do Tags Help Users Find Things?," *Knowledge Organization* 37, no. 4 (October 2010): 239-55; Yi-ling Lin et al., "The Impact of Image Descriptions on User Tagging Behavior: A Study of the Nature and Functionality of Crowdsourced Tags," *Journal of the Association for Information Science & Technology* 66, no. 9 (September 2015): 1785-98; Marija Petek, "Comparing User-generated and Librarian-generated Metadata on Digital Images," *OCLC Systems & Services: International Digital Library Perspectives* 28, no. 2 (May 25, 2012): 101-11; Katherine Rawson, "Curating Menus: Digesting Data for Critical Humanistic Inquiry," in *Laying the Foundation*, vol. 7, Digital Humanities in Academic Libraries (Purdue University Press, 2016), 59-72; Parul Zaveri and Mukta Atkekar, "Collaborative Tagging in Digital Libraries," *International Journal of Information Dissemination & Technology* 4, no. 2 (April 2014): 148-54.

¹⁹ Ari-Veikko Anttiroiko and Reijo Savolainen, "Towards Library 2.0: The Adoption of Web 2.0 Technologies in Public Libraries," *Libri: International Journal of Libraries & Information Services* 61, no. 2 (June 2011): 87-99; Matt Enis, "Wisdom of the Crowd," *Library Journal* 140, no. 11 (June 15, 2015): 38-40;

-
- Carolyn Foote, "Trending Now ... And Into the Future," *Internet@Schools* 24, no. 5 (December 11, 2017): 10–12; Peter J. Rolla, "User Tags versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?," *Library Resources & Technical Services* 53, no. 3 (July 2009): 174–84; Parul Zaveri and Mukta Atkekar, "Collaborative Tagging in Digital Libraries," *International Journal of Information Dissemination & Technology* 4, no. 2 (April 2014): 148–54.
- ²⁰ Shari Laster et al., "Transitioning from Marketing-Oriented Design to User-Oriented Design: A Case Study," *Journal of Web Librarianship* 5, no. 4 (October 2011): 299–321.
- ²¹ Heather Dawson, "Crowdsourcing the Library and Archive," *Multimedia Information & Technology* 41, no. 4 (November 2015): 13–14; Max J. Evans, "Archives of the People, by the People, for the People," *The American Archivist* 70, no. 2 (2007): 387–400..
- ²² Mary Flanagan and Peter Carini, "How Games Can Help Us Access and Understand Archival Images," *The American Archivist* 75, no. 2 (2012): 514–37; Dinah Sanders, "Tag – You’re It!," *American Libraries* 39, no. 11 (2008): 52–54; Jessica Short, "Take Ten to Tag! Smithsonian Gardens Public Tagging Initiative," *Technical Services Quarterly* 31, no. 4 (October 2, 2014): 319–31; "Metadata Games - ZenTag," accessed March 21, 2018, <https://tiltfactor2.dartmouth.edu/zentag>.
- ²³ Heather Dawson, "Crowdsourcing the Library and Archive," *Multimedia Information & Technology* 41, no. 4 (November 2015): 13–14; Mary Flanagan and Peter Carini, "How Games Can Help Us Access and Understand Archival Images," *The American Archivist* 75, no. 2 (2012): 514–37.
- ²⁴ Liliana Melgar Estrada et al., "Time-Based Tags for Fiction Movies: Comparing Experts to Novices Using a Video Labeling Game," *Journal of the Association for Information Science & Technology* 68, no. 2 (February 2017): 348–64.
- ²⁵ Parul Zaveri and Mukta Atkekar, "Collaborative Tagging in Digital Libraries," *International Journal of Information Dissemination & Technology* 4, no. 2 (April 2014): 148–54.
- ²⁶ Heather Dawson, "Crowdsourcing the Library and Archive," *Multimedia Information & Technology* 41, no. 4 (November 2015): 13–14.
- ²⁷ Gunho Chae et al., "Linking and Clustering Artworks Using Social Tags: Revitalizing Crowd-Sourced Information on Cultural Collections," *Journal of the Association for Information Science & Technology* 67, no. 4 (April 2016): 885–99; J. Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept," *New Review of Hypermedia and Multimedia* 12, no. 1 (June 1, 2006): 83–105; Linda Zajac, "Social Metadata Use in Art Museums: The Case of Social Tagging," *PNLA Quarterly* 77, no. 2 (Winter 2013): 65–77.
- ²⁸ Lesley L. Parilla and Meghan Ferriter, "Social Media and Crowdsourced Transcription of Historical Materials at the Smithsonian Institution: Methods for Strengthening Community Engagement and Its Tie to Transcription Output," *The American Archivist* 79, no. 2 (2016): 438–460; Jessica Short, "Take Ten to Tag! Smithsonian Gardens Public Tagging Initiative," *Technical Services Quarterly* 31, no. 4 (October 2, 2014): 319–31.
- ²⁹ Kais Dukes, Eric Atwell, and Nizar Habash, "Supervised Collaboration for Syntactic Annotation of Quranic Arabic," *Language Resources and Evaluation* 47, no. 1 (2013): 33–62.
- ³⁰ Martin Moyle, "Manuscript Transcription by Crowdsourcing: Transcribe Bentham," *Liber Quarterly: The Journal of European Research Libraries* 20, no. 3/4 (January 2011): 347–56.
- ³¹ "Citizen Archivist Missions," National Archives, May 15, 2017, <https://www.archives.gov/citizen-archivist/missions>.
- ³² Jan Zastrow, "The Digital Archivist. Crowdsourcing Cultural Heritage: 'Citizen Archivists' for the Future," *Computers in Libraries* 34, no. 8 (October 2014): 21–23.
- ³³ Tarvo Kärberg and Koit Saarevet, "Transforming User Knowledge into Archival Knowledge," *D-Lib Magazine* 22, no. 3/4 (April 3, 2016).
- ³⁴ Janelle Varin, "iTunes Metadata and Classical Music: Issues and Solutions for Crowdsourced Metadata in iTunes," *Serials Librarian* 69, no. 1 (July 2015): 70–76.

-
- ³⁵ Liliana Melgar Estrada et al., "Time-Based Tags for Fiction Movies: Comparing Experts to Novices Using a Video Labeling Game," *Journal of the Association for Information Science & Technology* 68, no. 2 (February 2017): 348–64.
- ³⁶ Pamela H. Mayer, "Like a Box of Chocolates: A Case Study of User-Contributed Content at Footnote," *The American Archivist* 76, no. 1 (2013): 19–46.
- ³⁷ Edward Benoit, "#MPLP: A Comparison of Domain Novice and Expert User-Generated Tags in a Minimally Processed Digital Archive" (The University of Wisconsin - Milwaukee, 2014); Dinah Sanders, "Tag – You're It!," *American Libraries* 39, no. 11 (2008): 52–54.
- ³⁸ Ciro Cattuto, Vittorio Loreto, and Luciano Pietronero, "Semiotic Dynamics and Collaborative Tagging," *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 5 (2007): 1461–1464; Gunho Chae et al., "Linking and Clustering Artworks Using Social Tags: Revitalizing Crowd-Sourced Information on Cultural Collections," *Journal of the Association for Information Science & Technology* 67, no. 4 (April 2016): 885–99; Youngok Choi and Sue Yeon Syn, "Characteristics of Tagging Behavior in Digitized Humanities Online Collections," *Journal of the Association for Information Science & Technology* 67, no. 5 (May 2016): 1089–1104; Max J. Evans, "Archives of the People, by the People, for the People," *The American Archivist* 70, no. 2 (2007): 387–400; Dinah Sanders, "Tag – You're It!," *American Libraries* 39, no. 11 (2008): 52–54.
- ³⁹ Elaine Ménard, "Image Retrieval: A Comparative Study on the Influence of Indexing Vocabularies," *Knowledge Organization* 36, no. 4 (October 2009): 200–213.
- ⁴⁰ Ellen Daly and Neil Ballantyne, "Ensuring the Discoverability of Digital Images for Social Work Education: An Online 'Tagging' Survey to Test Controlled Vocabularies," *Webology* 6, no. 2 (June 2009): 1–16; Marija Petek, "Comparing User-generated and Librarian-generated Metadata on Digital Images," *OCLC Systems & Services: International Digital Library Perspectives* 28, no. 2 (May 25, 2012): 101–11; J. Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept," *New Review of Hypermedia and Multimedia* 12, no. 1 (June 1, 2006): 83–105.
- ⁴¹ Corinne Jörgensen, Besiki Stvilia, and Shuheng Wu, "Assessing the Relationships Among Tag Syntax, Semantics, and Perceived Usefulness," *Journal of the Association for Information Science & Technology* 65, no. 4 (April 2014): 836–49.

Methodology

Overview

This study is a qualitative study exploring two different approaches to volunteer-based tagging in archival collections. In order to determine how controlled vocabularies and folksonomies applied in crowdsourcing initiatives affect discoverability of archival materials, an interview protocol will be used. First, several different institutions currently using crowdsourcing technologies will be selected for study. These institutions will represent collections using controlled vocabularies, those using folksonomies or uncontrolled tagging, and those using a combination of both. Archivists involved with these projects will be interviewed to gain a deeper understanding of how and why these methods are utilized. These interviews will give the professional archivist perspective on the issues at hand. This study aims to give insight into this new realm of public participation in archives.

Sampling

In order to explore how controlled vocabularies vs folksonomies affect discoverability in crowdsourced archival collections, archivists who have worked on these projects were interviewed. With the aim of exploring how different institutions view the issues at hand, interviewees were selected for crowdsourcing projects and willingness to participate in the study. Due to the specificity of the research need, institutions were selected using purposive sampling methods. In addition to this,

snowball sampling was used when contacts at the chosen institutions had suggestions of other archivists who would like to participate. Ideally, this study would have interviewed archivists involved with projects at four separate institutions, with no more than two participants at each institution. This gave an upper limit of eight participants for this stage of data collection.⁴² Out of the five institutions selected, only four participated. Only one of the four institutions had more than one archivist participate in the interview, bringing the total number of participants to 5 archivists/library professionals.

Email served as the primary communication method throughout sampling. Repositories were chosen and sent an email explaining the study and asking for participation.⁴³ Each project or repository contacted about participating in the study were sent the same email asking for participation.⁴⁴ Out of the five repositories contacted, four responded and agreed to participate in the study. One interviewee asked for additional information before agreeing to participate, but none refused participation after responding. Only one repository contacted did not respond to the request for participation, giving a response rate of 80%.

Data Collection Strategy

This study used semi-structured interviews in order to collect data. Interviews were conducted via phone, with the option of email if scheduling conflicts occurred. If a repository was close enough, participants were given the option of an in-person interview. In the end, all four institutions participated via phone interviews. The interviews were recorded with the permission of the person being interviewed and were transcribed at a later date by the author. Each participant had the option to break the

discussion into several interviews, on a case by case basis. This proved unnecessary and each of the four interviews were conducted in one sitting.⁴⁵

Data Analysis

The information from the interviews was transcribed and combed through for themes and similar thoughts throughout. Thematic analysis is a technique often used with exploratory qualitative data. In this technique, conversations, interviews, and other observational techniques are recorded, transcribed, and then coded for themes and recurring ideas. To ensure the integrity of the coded content, the researcher must be sure to use terms consistently and define what the terms mean and how to recognize them in the transcripts. It is also prudent to double check coded materials at least once after all transcripts have been coded to ensure cross-continuity. A key component of this theory of data analysis is that the researcher let the data inform the code. The researcher should not go into the data collection or the analysis phase with any preconceived notions of what they will find.⁴⁶ Because of this, what was coded for and discovered in this first phase of analysis was not pre-determined.

Implications

The purpose of this study was to explore whether controlled vocabularies or folksonomies are better used for tags in crowdsourced archival collections in terms of discoverability. The results of the research can be generalized and applied in archives and cultural institutions using crowdsourcing to tag their collections. They can look at the pros and the cons for both controlled vocabulary and folksonomy use in these projects and decide which suits their collections and needs. Having this information will significantly reduce the amount of backlogged unprocessed materials in archives, as

validating and streamlining crowdsourcing means less time and resources need to be spent processing collections.

⁴² Barbara M. Wildemuth, "Chapter 15: Sampling for Intensive Studies," in *Applications of Social Research Methods to Questions in Information and Library Science, 2nd Edition* (ABC-CLIO, 2016): 136-144.

⁴³ See Appendix A

⁴⁴ See Appendix B

⁴⁵ Barbara M. Wildemuth, *Applications of Social Research Methods to Questions in Information and Library Science, 2nd Edition* (ABC-CLIO, 2016): 239-257.

⁴⁶ Michi Komori, "Thematic Analysis," *Design Research Techniques*, <http://designresearchtechniques.com/casestudies/thematic-analysis/>.

Findings

After all four interviews were completed, the transcripts were reviewed and key themes, concepts, and ideas were identified. Although not all of the interviewees agreed on every point, a trend did begin to emerge. The professionals gave their views and opinions on crowdsourcing and the effectiveness of controlled vs uncontrolled vocabulary through the lenses of their own projects and personal experiences. Because, of this, a brief description of each participant and their project will be given before diving into their answers. These descriptions will provide enough background for one to understand their position within the field, but will not go into specifics which would jeopardize their anonymity.

Participant 1 is a systems librarian in a university system. They are a recent graduate with an MLS and began working on their crowdsourcing project as a graduate student. Their crowdsourcing project (identified as Iconography throughout this paper) is unique in that they are attempting to use crowdsourcing to build a thesaurus of terms, or a controlled vocabulary. This thesaurus will then be used by professionals in the field to apply metadata to museum objects. Volunteers in Iconography look at museum objects and identify the iconography contained in them using provided terms, but they can also add their own. In this sense, Iconography utilizes both controlled vocabulary and uncontrolled tags in the hopes of eventually integrating those tags into a larger, more comprehensive thesaurus.

Participant 2 is the outreach and communications manager at their institution. They have a background in studio design in addition to an MLS. This crowdsourcing project (identified as Illustrations) encompasses many different institutions and collections and has spanned various platforms over the years. The overarching Illustrations project has had four main initiatives. The first project initiative asks volunteers to identify species names and other pertinent information found in scientific illustrations using Flickr. The second project is similar in design to the Flickr project, but is hosted on the Zooniverse platform. Volunteers are once again asked to tag scientific illustrations and are given specific instructions on how to format these tags. Although users are encouraged to use standardized language provided in the instructions and as outlined in taxonomic species naming, they have the ability to vary from this structure to add in additional information. The last two projects involve transcription, and the specifics are therefore not relevant to this paper.

Participant 3 is the coordinator of a large-scale institutional transcription project (Transcription). They have only been in this position for 6 months but were involved on the other side of the project before transitioning to this new role. This gives the participant a unique perspective as both the coordinator of Transcription and as a professional user/client of the project as well; therefore, they gave insight into how the project works and also how well the information integrated back into the individual collections' databases. Project Transcription unfortunately does not currently use tagging, instead only focusing on transcription to make collections more accessible. Despite this, the information gathered during this interview was insightful for the impact of crowdsourcing projects as a whole and in deciding how and when to integrate tagging functionalities into these projects.

Participants 4 and 5 are both community managers for their institution's archives. Their institution is large, and their collections encompass many different organizations, time periods, and mediums. The vast nature of their collections makes their experiences unique, as they believe it would be difficult to ever create a controlled vocabulary to encapsulate all relevant information present in their holdings. Their project (Collections) is a crowdsourcing tagging project using uncontrolled vocabulary. Volunteers are encouraged to tag collections with relevant information but are also encouraged to use tags for personal reasons. Some users create unique identifiers for their classes, so that all the records they want to use in a teaching lesson will come up with one easy search. The project also features comment spaces, where users can comment on collections, talk to each other, and tell personal stories. Out of the 4 projects, this one utilizes the most folksonomic/uncontrolled approach.

These projects provide an accurate cross divide of the multitude of different crowdsourcing initiatives being used in cultural heritage institutions in terms of size, scope, and technology platforms. The one area of crowdsourced metadata applications that is not represented in this study is tagging and transcription games. Although these games were successful in pulling in users, they have proved less valuable in providing accurate and pertinent data. Gaming formats also do not lend themselves to volunteer collaboration and proofing.⁴⁷

Although each of the four projects are unique from one another, several reoccurring themes were found. The participants discussed time, volunteer engagement, value added, credibility, accessibility, and of course, controlled vs uncontrolled vocabulary.

Time

One of the most common motivations for crowdsourcing initiatives in cultural heritage institutions is the time it saves professionals. Adding item level metadata, whether it be descriptions, tags, or transcriptions, is extremely time consuming and often simply not feasible for an institution to commit to using only staff time and labor. The participants at Collections pointed this out, saying that their project “was developed as a way to help access our records even better. And the only way that it was going to be possible was through crowdsourcing, because archivists, especially archivists at [Collections]... [would not be able to] with the size and the number of records we have.” Despite the feasibility issues present in item level description, professionals do see the value in making specific item level information available to researchers. Discoverability increases when items are described individually in the online system. This is where crowdsourcing comes in. Professionals know that they do not have the resources to complete an item level description project, so they ask for the help of interested volunteers to complete the task for them. At Collections, the archivists explained, saying “It's done on such a big scale that you can't read every page and see, you wouldn't necessarily know what goes into every single page on the record. So, by having a whole cohort of [volunteers] who were willing to transcribe, they're unlocking those stories within the records. So, we think all of that helps with discoverability.” In addition to aiding discoverability for researchers, this crowdsourcing also positively affects the professionals at the involved institutions. Time is saved not only in the application of metadata, but also through the increased discoverability of the collections once the metadata is made searchable. The archivists and curators at Transcription have experienced this first-hand. Participant 3 explained this, saying “the curators are planning a new exhibit on World War I and wanted to include some sort of quotes and

other information from [the diary of air force pilot] Brooks, but it would take a ton of time for them to sit there and read everything versus getting it transcribed. Our volunteers got it transcribed in under a week, and now the curators can just search through that." The time that this saves the librarians, archivists, and museum professionals is immeasurable, as is the value added to the collections.

Volunteer Engagement

While time can be saved when volunteers help to process collections, this is only successful when volunteers are willing to do the work. Because of this, continuing user engagement is critical to a crowdsourcing initiative. Volunteer engagement includes thoughts on how to invite volunteers to participate in an initiative, how to keep them interested in the projects and coming back, and how to design systems that are easy to use and do not hinder volunteer participation. The participants at collections discussed the importance of this in depth, saying that "this is also part of a long-term strategic plan and our strategic goals for the [institution] to engage the public with American history and to connect them with our mission and encourage this crowdsourcing or encouraging [volunteers] to share their knowledge and become more engaged with the record." As can be seen here, volunteers are seen as the assets they are and keeping them engaged in order to continue harvesting their knowledge is paramount. Making sure that volunteers know about the project is just the first step, "sometimes... [the volunteers] need some encouragement, some hints and some inspiration," according to the participants at Collections. This is just as important as gathering volunteers, because in order to do good work, volunteers need instructions and inspiration. The participant at Iconography agreed with this notion of volunteer engagement and structure. Despite the clear instructions for their crowdsourcing project, the volunteers were only willing to

participate if they could input the data they wanted to contribute. As Participant 1 explains, “they were coming from a very specific domain knowledge region and when they were participating, they made a lot of use of the sort of ‘what additional terms would you use’ field and used that free text to go on what you might think of as more conventional object description. Doing long text description of the resource, the object, both the object type, the object style, and the iconographic content of it. So, while we did see that the terms we provided were useful in so much as selecting from menus, there were a lot of people that wanted to take advantage of things that were more familiar to them in their practice which was less structured more free text.” In this case, volunteers were given the opportunity to use this free text feature as it not only provided additional, potentially useful, information, but it also kept volunteers returning to the project. They felt their contributions were being heard. Concerns with controlled vocabularies vs uncontrolled tags and the amount of data harvested are at the forefront of this discussion in the answers from the participants, as can be seen in the controlled vs uncontrolled discussion.

Value Added

One aspect of these interviews that was unexpected was the focus on the value of volunteer metadata that cannot be replicated by the professionals at the institutions. The participants interviewed all agreed that volunteers can contribute so much more than just time and manpower to these projects. Participant 1 noted that “the enhancement value that you get from crowdsourcing is undeniable.” Oftentimes, volunteers have more knowledge about the collections than the archivist and cultural heritage professionals do. As stated by Participant 3, “The volunteers are way better at this than us at this point.” The participants from Collections noted that the metadata

coded by volunteers “really helps kind of democratize the records and helps really tell the story better than anybody else could.”

Subject experts, retired scholars, and individuals who were present at events, or have heard stories from those who were, can give our collections more context and deeper, richer information than would be possible without them. Transcription has really noticed this, saying “we have volunteers who tell us...they were professional proofreaders and volunteers who have been librarians and archivists and museum professionals...So it’s all of that information, all of that background, all of those skill sets really helps, especially with when they’re working together and working with us, it helps ensure that these projects are done well and done correctly.” At Iconography, the different users tag using a variety of terms, which “aligns the realities of practice that exist in how users describe things from a domain specific perspective or from a personal subjective use perspective. Both of which are meaningful pieces of information... [using controlled vocabulary] would have changed the results, maybe perhaps giving us cleaner data, because we would have had less variants and less opportunity for divergence between our terms and their terms, it wouldn’t have captured what the user expectations of use are.” Because volunteers have different perspectives, they often add metadata that others would not think of. While some professionals do not like that the metadata does not “match”, others see the value for discoverability and use that these deviations provide. Collections notes that “We're really trying to democratize the records and let people tell us what they know and what they see and their experience of it. So, these are their records, the records of the American people.”

The personal use of tags by volunteers in Collections is both a common occurrence in crowdsourcing and an unexpected added value to the project. Many

articles have identified and discussed the use of personal or subjective tags that are only of use to the person coding them. The personal tags in Collections are no different in that sense; only the people who applied these specific code tags find them useful. What is different in this project, however, is that users are encouraged to use the tagging feature in this way once the professional archivists saw others using it successfully in this way. The archivists at Collections did not develop their tagging feature with this type of use in mind, but when it became evident that some volunteers were tagging to make items easily retrievable for themselves, they decided to embrace the new and unexpected application of the technology. Participant 4 discussed this phenomenon, saying, "I don't think we expected people to kind of create a code like that, but then we discovered that a code of letters and numbers mixed together that someone wouldn't randomly use could make it unique and could make it easier to create a list...And then when we discovered how well they worked, we shared that as widely as we could in our workshops and lectures and things like that." Instead of simply ignoring this unconventional utilization of tagging or trying to steer volunteers away from using tags for their own benefit, the archivists shared this new knowledge with volunteers and encouraged them to use tags to create their own mini-collections within their vast repository.

Credibility

While volunteers certainly provide new insights and different ways of tagging collections, many professionals in the literature have expressed concerns regarding credibility and accuracy of crowdsourced metadata. Despite the prevalence of this in the literature, the five professionals interviewed said that these concerns are largely unfounded. All four projects have measures in place to review and correct the data

coded by volunteers; however, the professionals have discovered that volunteers go above and beyond time and time again to ensure that their contributions are as accurate as possible.

The specific measures for monitoring volunteer contributions vary institution to institution. The smaller project, Iconography, checks volunteer tags manually, with a professional on the team looking at new suggestions for keywords and deciding if they add value or if they are too close to existing vocabulary in the thesaurus. Because Iconography is such a niche project, volunteers usually self-select in or out based on their subject knowledge, so the accuracy of the thesaurus tags is less of a concern. This was noted by Participant 1, who said “we probably had a little more representation from the people coming from those programs as opposed to the general public because there is sort of a self-selecting and some of the other projects on [the site] are a bit more interesting than describing objects.” They did note, however, that “you’re also needing to do some sort of editorial work in trying to evaluate and integrate user supplied terminologies into your controlled authority. So that changes the work a little bit, but it does give you a plurality of perspectives when creating controlled terminologies which is something that has traditionally not been present in authority creation.”

The most challenging aspect of this project in terms of credibility was sorting through the uncontrolled tags and distilling the useful information to integrate into the controlled vocabulary. In regard to integration, “the crowdsourced terms themselves were not automatically integrated into our controlled vocabulary...these things go through an...editorial lens. So, if we had a suggestion of a term, we would review that term and think [through the questions determining if it should be integrated] That being said, when we do start the platform in actual production, there will be a tagging

element and that will be distinct from the iconographic description that we are applying.”

Illustrations has not yet created a system to integrate their crowdsourced collections back into the institutional records for the items. One of their partners, however, has begun importing crowdsourced data into their system. Because the data is not reviewed by other volunteers in any systemized way, this partner organization has created a verified user type of program, where volunteers can go through a vetting process and become trusted curators. Trusted curator volunteers can then go into the system and mark items as reviewed and a trusted symbol will appear in the record. Unreviewed information has an unreviewed symbol in the place of the trusted symbol. When planning the next steps for eventually ingesting the crowdsourced information back into the records, Illustrations is concerned with reviewing information and making sure that it has been coded according to current best practices, saying “we’ve actually just begun the process of updating our metadata model and one of things we are going to be considering is all of this crowdsourced metadata that we’ve gotten and where that fits within the model, how we display it in the library, how we indicate for example that this is data that came from crowdsourcing...initiatives and not for example from a curator.” This is a primary concern because they “have had people voice concerns over the tags and particularly because...one of the main tasks of taxonomy is determining what is the current valid scientific name and where do all the other names fall concerning that.” In particular, they “have had some concerns of people saying, ‘if it’s not a scientific expert within this particular species saying this is the current valid scientific name, then I don’t know if I can trust it.’” One way Illustrations is attempting to mitigate this concern is through encouraging “volunteers...to provide a link or

citation to the source that they used to determine that particular name, that particular attribution [in the comments section]. Despite their concerns about credibility and trusting volunteers' knowledge, they believe "if you structure your project well, provide adequate training, you can address a lot of those credibility concerns. You make sure you provide support for your users throughout the process and a lot of good documentation and resources they can refer to."

In Transcription, volunteers police and support each other through comments and review systems in place. Volunteers work together to transcribe pages and then review them for accuracy before marking them complete. Transcription also has an easy system to correct completed data if a mistake is found. As Participant 3 explained, "it's always automatically updated, so what we tell the volunteers and other is that nothing's ever set in stone. So, a project could be completed...but if a volunteer comes along and realizes that there's a typo or a mistake or a word missing, we can always reopen that project, they can edit, and then it automatically updates. So, projects are never completely done at least." In addition, individual units within the institution can review and approve transcriptions before importing the data; however, most of the professional participants have realized that this step is not needed and do not require professional approval before integrating the data. This process was implemented because "accuracy was sort of the biggest fear when transcription center was first developed in 2013, internally at least." This process is so rarely used, though because "the volunteers have just time and time again and only continued to impress and amaze us with how not only fast they are, but how dedicated and how meticulous they are."

Finally, Collections has no formal review system in place for its metadata. Instead, the user generated tags are shown as separate from the archival catalog records,

although both layers are searchable within the public facing catalog. As Collections explains, “the crowdsourced contributions are not part of the permanent record. It is a layer over our catalog that is then searched together with the catalog, but it is not then part of the permanent description that is then incorporated with what our professional archivists have created. It's separate, there's a back end that that's where that database is and then this is layered on top for the searching in the catalog.” This way, users can see which data was inputted by professionals and which tags were assigned by volunteers. Tags can only be removed by those who created them or by institution staff. This goes along with their emphasis on user control of collections. “We don't know what motivations are for people to tag things. John Smith may have challenged John Smith on that thing because it's his favorite record ever. And so, we don't know people's motivations. And again, we emphasized that these are user contributions and that we don't know what their motivations are for doing different things.” One small way that they attempt to maintain the integrity of the collection is through requiring users to register to tag. This is unique, as most of the other projects discussed allow for both named and anonymous volunteers. Despite concerns over credibility of crowdsourced metadata, Collections “have found that people are finding mistakes in the actual record,” and not in the crowdsourced data.

Accessibility

The common thread that weaves through all of these themes is accessibility. Crowdsourcing is used first and foremost as a way to increase accessibility to digitized collections. While the opinions of the professionals differed on certain points, they all agreed that the data added by volunteers increased the accessibility of the collections for researchers and the general public. Iconography believes that “the use of

crowdsourcing descriptions does promote findability because you get sort of two levels of description that you can't always get in typical description...so you're hitting populations expectations that you might not get if you were doing middle of the road description by getting the super generic and super specific." This again ties back to the idea of added value; volunteers often add different data than would otherwise be associated with an item. In Illustrations, transcriptions and tags on illustrations and other unique materials "will mean that they are searchable, discoverable, people will be able to search for terms, for species with those and discover a lot more information." Without the crowdsourced metadata, "if you actually wanted to find an illustration of an actual type of species, that can be really hard to do." Although Transcription is only using transcriptions and not yet implementing tagging, their work also makes sources discoverable. Objects that were once hidden because they "have millions of materials, individual pages and objects and items digitized...transcription center helps to unlock the hidden content within those digitized collections." This is key because despite being digitized, "what's findable online is what's available in that catalog record so if somebody is searching for something that's specifically contained in the title or the open content note or the tags that are connected to it or the controlled vocabulary, then they'll find it great," but if someone is searching for something more specific, they will often only be able to find it with the transcription. So then, transcriptions make text and small details text-searchable, volunteers tagging illustrations make photos findable, and item specific information added to collections allows users to easily search and discover new resources. In the end, as Collections explains, "we're looking at the question 'how do we help people find what they're looking for?'...crowdsourcing...is helping people discover the things that we wouldn't necessarily see...when staff are processing records."

Controlled vs Uncontrolled

With all of this in mind, the question remains, which type of tagging is more conducive to discoverability? While it would be wonderful to have a clear-cut answer, the reality is that it depends on a number of factors. This was the consensus of the participants, that there is no consensus.

In favor of uncontrolled vocabulary usage, Iconography found “that people also sort of did not want to be shoehorned into the specific ways of describing things.” If they decided to eliminate the free text option, Participant 1 believed that “it would be less representative. The result we would get in that test would have been less representative of real use.” Despite these insights, however, Participant 1 said, “as a librarian I would have to think that controlled vocabularies and other sorts of structured authorities do provide better access simply because it is a concerted effort on the part of the organizers to reduce redundancy and to reduce conflict in the spirit of promoting co-location of resources, which is something that the subjectivity and semantics of crowdsourcing doesn’t necessarily have in mind.” So, in a perfect world with perfect users and a perfect controlled vocabulary, using a thesaurus option for crowdsourcing would be the best option; however, in an imperfect world where current thesauri are not all encompassing and user input is valued and creates potentially better results, a mixed approach is preferable.

In response to this question of controlled vs uncontrolled, many professionals point to uniformity and ease of data integration and transfer. When controlled vocabularies are used, they are easily integrated into existing systems and metadata schemas. Illustrations explained this, saying “the controlled vocabularies really do help

with discoverability because it's easier for us to ingest and do something meaningful with controlled vocabulary...it's easier for us to be able to ingest the data and make it searchable for users when it is a controlled environment and it's a common ground for people to search for something within the library." Transcriptions agreed on this point, adding that "in terms of controlled vocabulary, I think that any time we can sort of standardize and start to think about metadata on a large scale together across these institutions, it makes all of this easier and more accessible."

Though controlled vocabulary can make data transfer easier and more uniform, sometimes it simply is not a viable option. Participant 3 discussed how different projects require different approaches, saying "it really depends on the project because some things you're not going to be able to do controlled vocabulary for, some things you're not going to be able to do standardized metadata in general for, and it would just be easier to have sort of tags that they create." The participants at Collections agreed with this sentiment, explaining that they "have so many different types and kinds of records. And so, you don't always find the same things in every record," making it difficult to create a universal controlled vocabulary for their collections. They even went farther than this, saying that even if you had a smaller collection, "that a controlled vocabulary would be very useful, but...I feel that controlled vocabularies...are more for the librarian and the archival professional archives professionals and not the general public."

Instead of using controlled vocabularies, the participants at Collections "wanted to harness the wisdom of the crowd and see what people found in the records." This ties back to how they feel about volunteer contributions and the value they bring. In order to foster free-knowledge sharing, they "didn't want to dictate what we thought

they should see or find. We wanted to see what they were going to discover on their own.” Transcription agreed on this point, adding that discoverability improves with more information. Participant 3 stated “and I think any additional thing like that, when you’re talking about tags in general are really useful just to create different access points because everyone searches differently, so I think the more the better.”

⁴⁷ Liliana Melgar Estrada et al., “Time-Based Tags for Fiction Movies: Comparing Experts to Novices Using a Video Labeling Game,” *Journal of the Association for Information Science & Technology* 68, no. 2 (February 2017): 348–64.

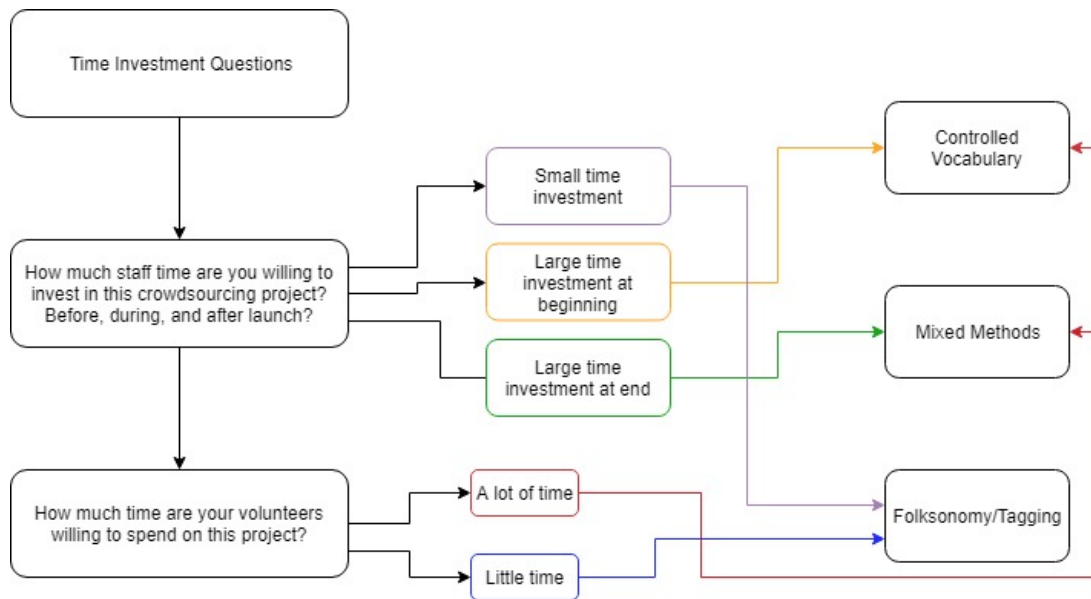
Discussion

So, what does all of this mean for crowdsourcing and discoverability?

Crowdsourcing approaches and successes rely on the collections themselves. Although there is no one correct way to apply metadata to collections in a crowdsourcing initiative to best promote discoverability, different methods can be more successful based on each unique institution and their collections. Throughout the conversations with professionals in the field, they all intimated that their answers depended on the specifics of the collection in question. For some collections, it makes more sense to use a controlled vocabulary and add more structure to the items, but for others it might be more valuable to use a folksonomy and gain new and unique information about the materials. Deciding when to use each approach is key to providing access to collections.

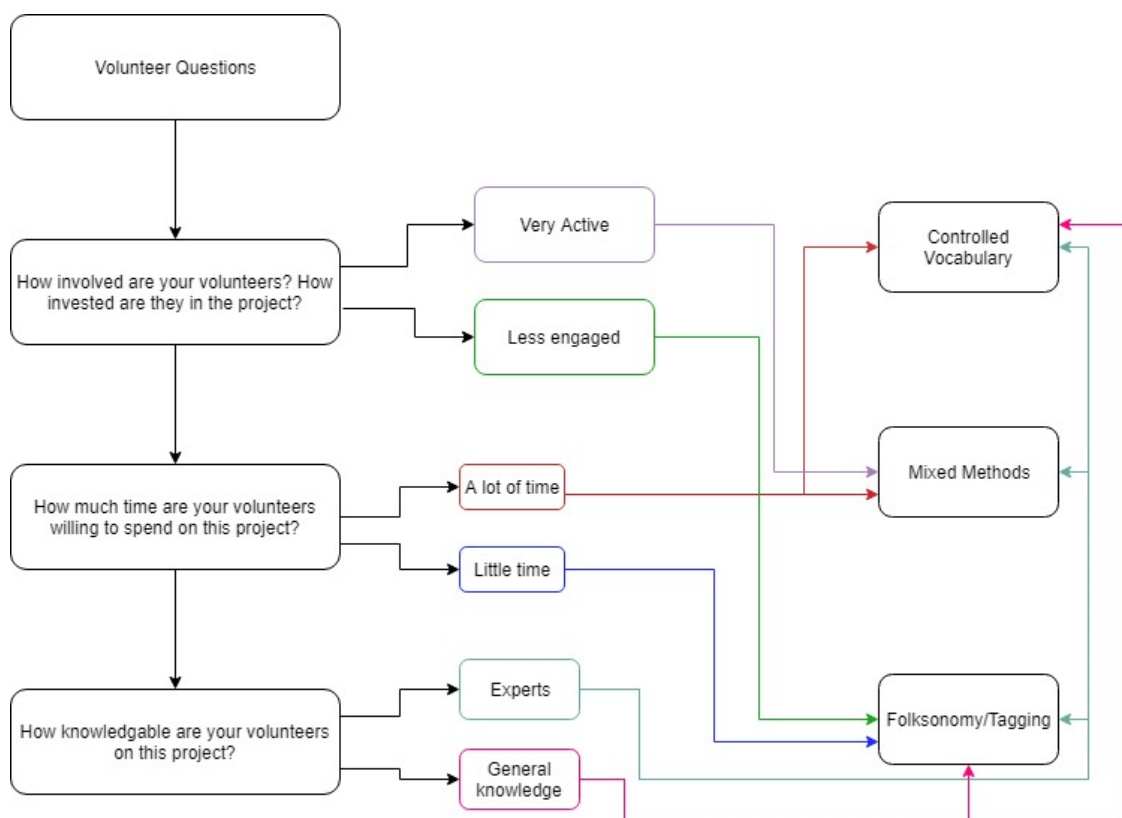
This decision will be based on multiple questions. Examples of the types of questions archivists should ask themselves are: Why are you digitizing the collection? How will users interact with the finished product? What information will end users be searching for? Do you have a dedicated volunteer base? How large is the collection? Looking at all of the questions and their answers will help professionals come to a conclusion for the best course of action for their institutional needs. By looking at the factors of time, volunteer engagement, value added, and accessibility, individual projects can evaluate which approach is best for them.

Controlled vocabularies and thesauri can be wonderful tools, guiding volunteers in tagging and making sure they are identifying the correct (or wanted) themes in the digitized content. The time invested in creating a controlled vocabulary, however, can be inhibitive depending on the size and scope of a collection. Institutions interested in crowdsourcing should determine how much professional time they wish to spend managing the crowdsourcing project versus how much time they want their volunteers to commit. Even mixed approaches, such as that utilized by Iconography, can be time intensive on the part of the professionals involved in the project. The amount of time involved in crowdsourcing projects can differ based on methods, but most often there will be a significant time commitment on both the institution and the volunteers. The key is deciding at what point you want to carve out time and resources to work on the project. Frontloading the work by creating a controlled vocabulary means less time spent at the end integrating the data into a system, but means that resources are being used before knowing the success or failure potential of the project. Waiting until the end to try and integrate data using a mixed approach means little time spent in the beginning, but intensive screening and editing in the end phases. While professionals tend to think about institutional resources, it is equally important to think about volunteer resources. If volunteers have time to learn how to use more complex systems, a mixed methods or controlled vocabulary approach may work, but if they have little time, a simple uncontrolled tagging structure will work better. A method will only work and promote discoverability if volunteers and staff put in the time to make them work.



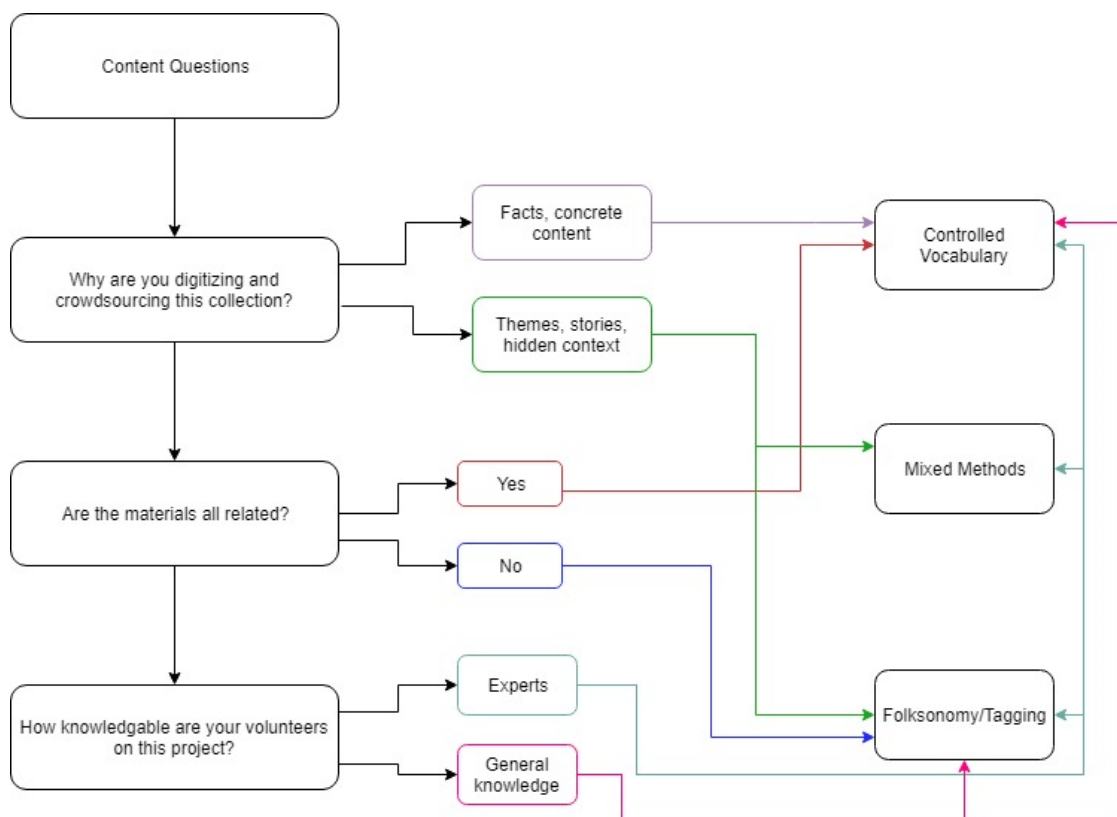
Volunteer engagement and their willingness to participate using various platforms and systems can also inform the direction of crowdsourcing. Illustrations noted that there can be a strong learning curve in the beginning of a project, especially when using a complex controlled vocabulary system, but that dedicated volunteers will persevere. Similarly, Iconography noted the need of their participants to be heard through the free text option. Knowing how volunteers want to communicate and participate can help decide the characteristics of tagging schemas. If volunteers are active and are willing and excited to spend time working on a project, then mixed methods may be the correct choice. This way, they can go through the controlled vocabulary options and choose the perfect tags, but can also elaborate on objects and give more detailed information. If they are less engaged, a simple uncontrolled tagging method may be best, so they do not have to learn how to use a complicated system. Time once again becomes a factor here, as their time corresponds closely to their participation levels. The base knowledge on a collection that volunteers bring with them can also inform the type of tagging used. Experts may want to give more specific

information that may add value to the materials. Less knowledgeable users may benefit from controlled vocabulary options guiding their tagging, or they might want to be able to tag in more general layman's terms using uncontrolled tags.

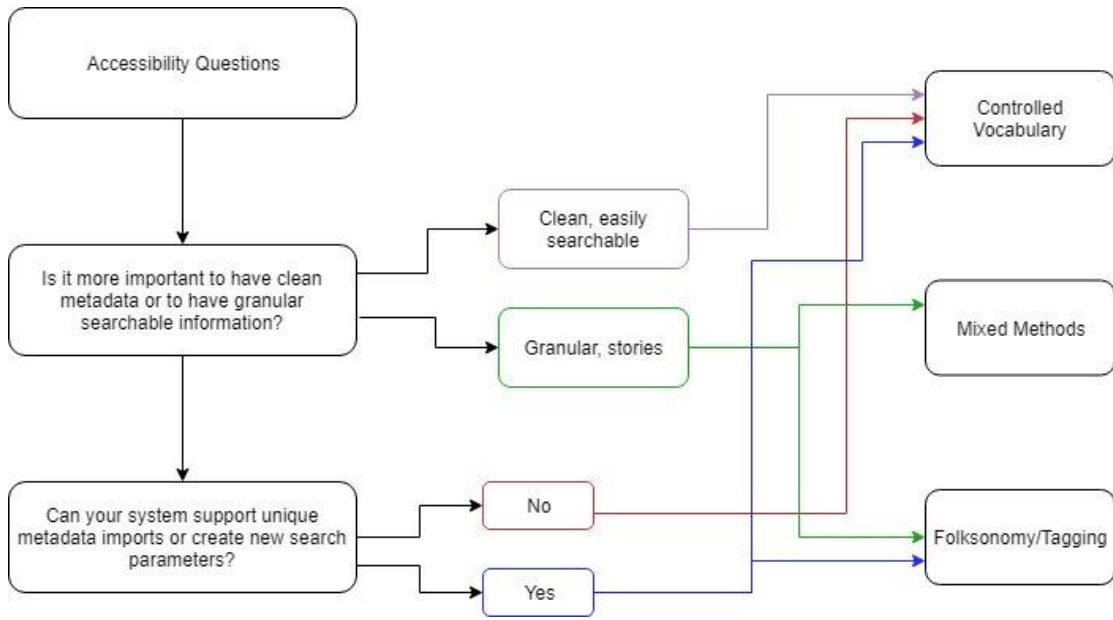


Deciding what type of information is hoped to gain through crowdsourcing will greatly impact which method is chosen. Volunteers have invaluable knowledge that can be integrated into item records through comments and tagging. If the records have more of a story to tell, using a less controlled system to encourage this dialog is preferable. If the records have less of a story and need more straightforward information coded, then a controlled vocabulary will help with authority control and minimizing synonyms to promote greater discoverability. If the materials are all related, this will make creating a controlled vocabulary easier. If, however, the collections are

broad and cover many different time periods, organizations, and themes, a controlled vocabulary may not be possible.



Accessibility is the reason that institutions decide to crowdsource metadata for their collections, but the ways in which they want users to access information can affect tagging choices. Again, the more uniform the information is, the easier it is to find all records on a particular subject, but some smaller details or more personal stories may be lost. The capabilities of the database and searching systems can also affect decisions on vocabularies as accessibility is only possible if the metadata can be integrated and searched in the systems. If the institution's systems can promote discoverability in collections with various metadata layers, then a mixed methods or uncontrolled tagging option will work. If not, then a controlled vocabulary that can be easily integrating into existing structures should be used.



Conclusions

The primary duty of an archivist is to protect, preserve, and make accessible the materials housed in their collections. While protection and preservation are well defined in archival training and practice, accessibility and use of collections is a more complex responsibility, which defies concise definitions.

As the amount of records entering archives continues to increase, the less time and resources archivists can devote to each item. Less Process, More Product can help quickly process materials and make them semi-available for use but can also inadvertently hide items in collections and hinder their discoverability; however, not processing collections and allowing backups to occur is even more detrimental for discoverability and access. Through crowdsourcing, materials in our institutions can be processed at a specific item level in a way that is simply not possible using only institutional staff.

In addition to providing the much-needed description for discoverability in our digital catalogs, crowdsourcing can also add value to collections because it utilizes the knowledge and experiences of so many individuals. While the information coded may not match the information an archivist would apply, it still aids in searching and can also provide new insights into the collections. Our society has a great deal to offer our archival collections, if we let them. Whether through controlled vocabularies or through

folksonomies, crowdsourcing volunteers never cease to amaze with the new and innovative ways they code data.

Appendix A: Email Form

Dear (Participant name),

Hello, my name is Ayla Toussaint and I am a graduate student at the University of North Carolina in the School of Information and Library Science. I am currently working on my Master's Paper, "Finding a Way Through the Crowd: How Keyword Choices Affect Discoverability in Crowdsourced Archival Tagging". This paper explores how using controlled vocabularies vs. folksonomies or uncontrolled vocabularies affect discoverability in crowdsourced collections. In order to do this, I am interviewing archivists who have personal experience with crowdsourcing initiatives.

I was hoping that you (or someone at your institution) would be willing to participate in my study. The interview can be conducted in person or via telephone and should last between 1-2 hours. Timing and scheduling are flexible based on your schedule! I would truly appreciate any information you could give me on this topic, especially as crowdsourcing becomes more popular and pressure to upload collections to the web increases.

Please find a consent form with more information below. If you have any other questions, please do not hesitate to ask! I look forward to hearing from you. I am also open to suggestions of other people to contact.

Sincerely,

Ayla Toussaint

Appendix B: Interview Question Template⁴⁸

Background Questions

1. Are you okay if I record this interview for my records?
2. Are you comfortable being quoted or referenced in my Master's Paper?
3. How long have you been an archivist/job title at X?
4. How did you end up in this position?/How did you become an archivist/job title?

Grounding Questions

5. Have you ever worked on a crowdsourcing project? Can you tell me about it?
6. What roles have you played in crowdsourcing projects?
7. Do you work closely with your volunteers?/Do you personally work with the volunteers or are they more anonymous?
8. Have you ever worked to build a controlled vocabulary for a collection?
 - a. Can you tell me about that process?

Specific Questions

9. In your crowdsourcing experience, have you used controlled vocabularies or just let volunteers create their own tags?

10. If you've used controlled vocabularies, do you think volunteers like the structure or find it cumbersome? If you haven't, how do you think they would feel based on your experience?
11. How do you implement crowdsourced collection metadata into your online systems?
 - a. How do you search in them?
12. Do you think crowdsourced collections improve discoverability of pieces in collections?
13. Do you think controlled vocabulary or unstructured tags lead to better discoverability?
14. Have you received any feedback from end users about crowdsourced tags?
15. Do you think that there is a credibility concern with crowdsourced collections?
Do you think that controlled vocabularies vs. folksonomies/uncontrolled tags affect credibility?

Wrapping up Questions

16. Do you have any other insights or angles that we haven't covered?
17. Is there anything you've said in this interview that you aren't comfortable with me using for my paper?
18. Would you like me to send you a transcript for edit?

⁴⁸ These are questions to guide the conversation; however, more questions may arise or other avenues of discussion may come up, depending on the interviewee. This will be a semi-structured interview, so the wording of the questions may change slightly during the interview, but the meaning will remain the same.

Bibliography

- Anderson, Scott R., and Robert B. Allen. "Envisioning the Archival Commons." *The American Archivist* 72, no. 2 (2009): 383–400.
- Anttiroiko, Ari-Veikko, and Reijo Savolainen. "Towards Library 2.0: The Adoption of Web 2.0 Technologies in Public Libraries." *Libri: International Journal of Libraries & Information Services* 61, no. 2 (June 2011): 87–99. <https://doi.org/10.1515/libr.2011.008>.
- Barrington, Luke, Douglas Turnbull, and Gert Lanckriet. "Game-Powered Machine Learning." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 17 (2012): 6411–16.
- Bell, Suzanne S. *Librarian's Guide to Online Searching*. 4th ed. Libraries Unlimited, 2015. <https://www.abc-clio.com/LibrariesUnlimited/product.aspx?pc=A4596P>.
- Benoit, Edward. "#MPLP: A Comparison of Domain Novice and Expert User-Generated Tags in a Minimally Processed Digital Archive." Ph.D., The University of Wisconsin - Milwaukee, 2014. <https://search-proquest-com.libproxy.lib.unc.edu/docview/1625718144/abstract/B432DE2E82374360PQ/1>.
- Cattuto, Ciro, Vittorio Loreto, and Luciano Pietronero. "Semiotic Dynamics and Collaborative Tagging." *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 5 (2007): 1461–64.
- Chae, Gunho¹, Jaram¹ Park, Juyong¹ Park, Woon Seung² Yeo, and Chungkon¹ Shi. "Linking and Clustering Artworks Using Social Tags: Revitalizing Crowd-Sourced Information on

Cultural Collections." *Journal of the Association for Information Science & Technology* 67, no. 4 (April 2016): 885–99. <https://doi.org/10.1002/asi.23442>.

Choi, Youngok¹, and Sue Yeon¹ Syn. "Characteristics of Tagging Behavior in Digitized Humanities Online Collections." *Journal of the Association for Information Science & Technology* 67, no. 5 (May 2016): 1089–1104. <https://doi.org/10.1002/asi.23472>.

"Citizen Archivist Missions." National Archives, May 15, 2017.

<https://www.archives.gov/citizen-archivist/missions>.

Daly, Ellen¹, ellen.daly@iriss.org.uk, and Neil² Ballantyne neil.ballantyne@iriss.org.uk.

"Ensuring the Discoverability of Digital Images for Social Work Education: An Online 'Tagging' Survey to Test Controlled Vocabularies." *Webology* 6, no. 2 (June 2009): 1–16.

Dawson, Heather. "Crowdsourcing the Library and Archive." *Multimedia Information & Technology* 41, no. 4 (November 2015): 13–14.

Dukes, Kais, Eric Atwell, and Nizar Habash. "Supervised Collaboration for Syntactic Annotation of Quranic Arabic." *Language Resources and Evaluation* 47, no. 1 (2013): 33–62.

Enis, Matt. "Wisdom of the Crowd." *Library Journal* 140, no. 11 (June 15, 2015): 38–40.

Estrada, Liliana Melgar¹, Michiel² Hildebrand, Victor³ de Boer, and Jacco³ van Ossenbruggen. "Time-Based Tags for Fiction Movies: Comparing Experts to Novices Using a Video Labeling Game." *Journal of the Association for Information Science & Technology* 68, no. 2 (February 2017): 348–64. <https://doi.org/10.1002/asi.23656>.

Evans, Max J. "Archives of the People, by the People, for the People." *The American Archivist* 70, no. 2 (2007): 387–400.

Flanagan, Mary, and Peter Carini. "How Games Can Help Us Access and Understand Archival Images." *The American Archivist* 75, no. 2 (2012): 514–37.

Foote, Carolyn. "Trending Now ... And Into the Future." *Internet@Schools* 24, no. 5 (December 11, 2017): 10-12.

Greene, Mark, and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68, no. 2 (September 1, 2005): 208-63.
<https://doi.org/10.17723/aarc.68.2.c741823776k65863>.

Hayman, Sarah, and Nick Lothian. "Integrating User Tagging and Controlled Vocabularies for Australian Education Networks." *IFLA Conference Proceedings*, November 2007, 1-27.

Hynning, Victoria Van, Samantha Blickhan, Laura Trouille, and Chris Lintott. "Transforming Libraries and Archives through Crowdsourcing." *D-Lib Magazine* 23, no. 5/6 (May 2017).
<https://doi.org/10.1045/may2017-vanhynning>.

Jørgensen, Corinne¹, cjorgensen@fsu.edu, Besiki¹ Stvilia bstvilia@fsu.edu, and sw09f@my.fsu.edu Shuheng Wu¹. "Assessing the Relationships Among Tag Syntax, Semantics, and Perceived Usefulness." *Journal of the Association for Information Science & Technology* 65, no. 4 (April 2014): 836-49. <https://doi.org/10.1002/asi.23029>.

Kacunko, Slavko. "M.A.D.: Media Art Database(s) and the Challenges of Taste, Evaluation and Appraisal." *Leonardo* 42, no. 3 (2009): 245-50.

Kärberg, Tarvo¹, karberg@ut.eetarvo.kargerg@ra.ee, and Koit² Saarevet koit.saarevet@ra.ee. "Transforming User Knowledge into Archival Knowledge." *D-Lib Magazine* 22, no. 3/4 (April 3, 2016): 11-11. <https://doi.org/10.1045/march2016-karberg>.

Kipp, Margaret E. I., and D. Grant Campbell. "Searching with Tags: Do Tags Help Users Find Things?" *Knowledge Organization* 37, no. 4 (October 2010): 239-55.

Laster, Shari, Tammy Stitz, Frank J. Bove, and Casey Wise. "Transitioning from Marketing-Oriented Design to User-Oriented Design: A Case Study." *Journal of Web Librarianship* 5, no. 4 (October 2011): 299-321. <https://doi.org/10.1080/19322909.2011.623517>.

"LHT-11-2013-0150.Pdf." Accessed February 14, 2018.

<http://www.emeraldinsight.com.libproxy.lib.unc.edu/doi/pdfplus/10.1108/LHT-11-2013-0150>.

Lin, Yi-ling¹, Christoph² Trattner, Peter³ Brusilovsky, and Daqing³ He. "The Impact of Image Descriptions on User Tagging Behavior: A Study of the Nature and Functionality of Crowdsourced Tags." *Journal of the Association for Information Science & Technology* 66, no. 9 (September 2015): 1785–98. <https://doi.org/10.1002/asi.23292>.

Lonie, Emily. "Many Hands Make Light Work? Crowdsourcing Archival Description and Transcription." *Things I'm Fond of* (blog), June 20, 2016.

<http://www.thingsimfondsof.com/crowdsourcing-archival-transcription/>.

Mayer, Pamela H. "Like a Box of Chocolates: A Case Study of User-Contributed Content at Footnote." *The American Archivist* 76, no. 1 (2013): 19–46.

McCutcheon, Sevim. "Keyword vs Controlled Vocabulary Searching: The One with the Most Tools Wins." *Indexer* 27, no. 2 (June 2009): 62–65.

Ménard, Elaine. "Image Retrieval: A Comparative Study on the Influence of Indexing Vocabularies." *Knowledge Organization* 36, no. 4 (October 2009): 200–213.

"Metadata Games - ZenTag." Accessed March 21, 2018.

<https://tiltfactor2.dartmouth.edu/zentag>.

Moyle, Martin¹, m.moyle@ucl.ac.uk. "Manuscript Transcription by Crowdsourcing: Transcribe Bentham." *Liber Quarterly: The Journal of European Research Libraries* 20, no. 3/4 (January 2011): 347–56.

Parilla, Lesley L., and Meghan Ferriter. "Social Media and Crowdsourced Transcription of Historical Materials at the Smithsonian Institution: Methods for Strengthening

Community Engagement and Its Tie to Transcription Output,” 2016.

<http://repository.si.edu//handle/10088/32112>.

“Parilla_Lesley-20161101-212_234_aarc_79_02_10.Pdf.” Accessed February 14, 2018.

https://repository.si.edu/bitstream/handle/10088/32112/Parilla_Lesley-20161101-212_234_aarc_79_02_10.pdf.

Petek, Marija. “Comparing User-generated and Librarian-generated Metadata on Digital Images.” *OCLC Systems & Services: International Digital Library Perspectives* 28, no. 2 (May 25, 2012): 101–11. <https://doi.org/10.1108/10650751211236659>.

Preez, Madely du. “Taxonomies, Folksonomies, Ontologies: What Are They and How Do They Support Information Retrieval?” *The Indexer* 33 (January 1, 2015).

Rawson, Katherine. “Curating Menus:: Digesting Data for Critical Humanistic Inquiry.” In *Laying the Foundation*, 7:59–72. Digital Humanities in Academic Libraries. Purdue University Press, 2016. <https://doi.org/10.2307/j.ctt163t7kq.7>.

Rolla, Peter J. “User Tags versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?” *Library Resources & Technical Services* 53, no. 3 (July 2009): 174–84.

Sanders, Dinah. “Tag – You’re It!” *American Libraries* 39, no. 11 (2008): 52–54.

Short, Jessica. “Take Ten to Tag! Smithsonian Gardens Public Tagging Initiative.” *Technical Services Quarterly* 31, no. 4 (October 2, 2014): 319–31. <https://doi.org/10.1080/07317131.2014.943005>.

“Speaking Technically.” *American Libraries* 39, no. 7 (2008): 54–57.

“The Nature of Tags in a Knowledge Organization System of Primary Visual Resources: Journal of Library Metadata: Vol 17, No 1.” Accessed February 14, 2018.

<http://www.tandfonline.com.libproxy.lib.unc.edu/doi/full/10.1080/19386389.2017.1314131>.

Trant, J., and with the participants in the steve museum project. "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept." *New Review of Hypermedia and Multimedia* 12, no. 1 (June 1, 2006): 83–105.

<https://doi.org/10.1080/13614560600802940>.

Varin, Janelle1. "iTunes Metadata and Classical Music: Issues and Solutions for Crowdsourced Metadata in iTunes." *Serials Librarian* 69, no. 1 (July 2015): 70–76.

<https://doi.org/10.1080/0361526X.2015.1036196>.

Wang, Aobo, Cong Duy Vu Hoang, and Min-Yen Kan. "Perspectives on Crowdsourcing Annotations for Natural Language Processing." *Language Resources and Evaluation* 47, no. 1 (2013): 9–31.

Wildemuth, Barbara M. *Applications of Social Research Methods to Questions in Information and Library Science, 2nd Edition*. ABC-CLIO, 2016.

Woodward, Eddie. "METADATA FOR IMAGE COLLECTIONS." *American Libraries* 45, no. 6 (2014): 42–44.

Yates, JoAnne. *Control Through Communication: The Rise of System in American Management*. JHU Press, 1993.

Zajac, Linda1, lindazajac@shaw.ca. "Social Metadata Use in Art Museums: The Case of Social Tagging." *PNLA Quarterly* 77, no. 2 (Winter 2013): 65–77.

zastrow, jan, zastrow@hawaii.edu. "The Digital Archivist. Crowdsourcing Cultural Heritage: 'Citizen Archivists' for the Future." *Computers in Libraries* 34, no. 8 (October 2014): 21–23.

Zaveri, Parul1, parulzaveri2004@yahoo.co.in, and Mukta2 Atkekar. "Collaborative Tagging in Digital Libraries." *International Journal of Information Dissemination & Technology* 4, no. 2 (April 2014): 148-54.