

8-1-2014

#MPLP: a Comparison of Domain Novice and Expert User-generated Tags in a Minimally Processed Digital Archive

Edward A. Benoit III

University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>



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

Recommended Citation

Benoit III, Edward A., "#MPLP: a Comparison of Domain Novice and Expert User-generated Tags in a Minimally Processed Digital Archive" (2014). *Theses and Dissertations*. 550.
<https://dc.uwm.edu/etd/550>

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact open-access@uwm.edu.

#MPLP: A COMPARISON OF DOMAIN NOVICE AND EXPERT USER-
GENERATED TAGS IN A MINIMALLY PROCESSED DIGITAL ARCHIVE

by

Edward Benoit, III

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Information Studies

at

The University of Wisconsin-Milwaukee

August 2014

ABSTRACT

#MPLP: A COMPARISON OF DOMAIN NOVICE AND EXPERT USER-GENERATED TAGS IN A MINIMALLY PROCESSED DIGITAL ARCHIVE

by

Edward Benoit, III

The University of Wisconsin-Milwaukee, 2014
Under the Supervision of Professor Iris Xie

The high costs of creating and maintaining digital archives precluded many archives from providing users with digital content or increasing the amount of digitized materials.

Studies have shown users increasingly demand immediate online access to archival materials with detailed descriptions (access points). The adoption of minimal processing to digital archives limits the access points at the folder or series level rather than the item-level description users' desire. User-generated content such as tags, could supplement the minimally processed metadata, though users are reluctant to trust or use unmediated tags.

This dissertation project explores the potential for controlling/mediating the supplemental metadata from user-generated tags through inclusion of only expert domain user-

generated tags. The study was designed to answer three research questions with two parts each: 1(a) What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?, 1(b) Are there differences

between expert and novice users' opinions of the tagging experience and tag creation

considerations?, 2(a) In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed

digital archive?, 2(b) Does user knowledge affect the proportion of tags matching

unselected metadata in a minimally processed digital archive?, 3(a) In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?, and 3(b) Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?

The dissertation project was a mixed-methods, quasi-experimental design focused on tag generation within a sample minimally processed digital archive. The study used a sample collection of fifteen documents and fifteen photographs. Sixty participants divided into two groups (novices and experts) based on assessed prior knowledge of the sample collection's domain generated tags for fifteen documents and fifteen photographs (a minimum of one tag per object). Participants completed a pre-questionnaire identifying prior knowledge, and use of social tagging and archives. Additionally, participants provided their opinions regarding factors associated with tagging including the tagging experience and considerations while creating tags through structured and open-ended questions in a post-questionnaire.

An open-coding analysis of the created tags developed a coding scheme of six major categories and six subcategories. Application of the coding scheme categorized all generated tags. Additional descriptive statistics summarized the number of tags created by each domain group (expert, novice) for all objects and divided by format (photograph, document). T-tests and Chi-square tests explored the associations (and associative strengths) between domain knowledge and the number of tags created or types of tags created for all objects and divided by format. The subsequent analysis compared the tags with the metadata from the existing collection not displayed within the sample collection

participants used. Descriptive statistics summarized the proportion of tags matching unselected metadata and Chi-square tests analyzed the findings for associations with domain knowledge. Finally, the author extracted existing users' query terms from one month of server-log data and compared the generated-tags and unselected metadata. Descriptive statistics summarized the proportion of tags and unselected metadata matching query terms, and Chi-square tests analyzed the findings for associations with domain knowledge. Based on the findings, the author discussed the theoretical and practical implications of including social tags within a minimally processed digital archive.

© Copyright Edward Benoit, III

All Rights Reserved

To my mother and father,
for all of your support

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION.....	1
1.1 Adaptions of Minimal Processing to Digital Archives.....	4
1.2 Research Problem, Questions, and Hypotheses.....	10
1.3 Research Design.....	13
1.4 Significance of Dissertation Project.....	14
1.5 Summary of Dissertation	16
CHAPTER TWO: LITERATURE REVIEW.....	18
2.1 Defining Digital Archives.....	21
2.2 Digital Archives.....	31
2.3 Minimal Processing	43
2.4 Postmodernism and Archives	55
2.5 Participatory Archives and Archives 2.0	58
2.6 Social Tagging	69
2.7 Social Tagging in Digital Libraries.....	76
2.8 Social Tagging in Digital Archives.....	79
2.9 Tagging Issues and Limitations	81
2.10 Expert and Novice Users	84
2.11 Limitations of Literature.....	88

CHAPTER THREE: METHODOLOGY	90
3.1 Sample Collection.....	93
3.2 Sample Population	96
3.3 Participant Demographics.....	99
3.4 Data Collection Methods and Procedures.....	105
3.5 Pilot Study.....	108
3.6 Data Analysis	108
3.6.1 RQ1a— What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?	109
4.6.2 RQ1b— Are there differences between expert and novice users’ opinions of the tagging experience and tag creation considerations?	113
3.6.3 RQ2a— In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive? RQ2b— Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?	114
3.6.4 RQ3a— In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users’ search terms in a digital archive? RQ3b— Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?.....	116
3.7 Validity, Reliability, and Generalizability	117
CHAPTER FOUR: RESULTS	120

4.1 Research Question 1(a): What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?	123
4.1.1 Number of Tags Generated by Expert and Novice Participants	123
4.1.2 Types of Tags Generated by Expert and Novice Participants	129
4.1.3 Similarities and Differences of Expert and Novice Participants' Tags	145
4.2 Research Question 1(b): Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?.....	153
4.3 Research Question 2(a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?	158
4.4 Research Question 2(b): Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?	165
4.5 Research Question 3(a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?	167
4.6 Research Question 3(b): Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?.....	172
CHAPTER FIVE: DISCUSSION.....	174
5.1 Theoretical Implications	174
5.2 Practical Implications.....	178
5.3 Methodological Implications	185

5.3 Limitations	186
CHAPTER SIX: CONCLUSION	188
6.1 Future Research	191
6.2 Future Directions	192
Bibliography	195
APPENDIX A: SAMPLE COLLECTION GROUPINGS	234
APPENDIX B: PARTICIPANT KNOWLEDGE ASSESSMENT	238
APPENDIX C: PRE-QUESTIONNAIRE	240
APPENDIX D: POST-QUESTIONNAIRE	242

LIST OF FIGURES

Figure 2.1:	Minimal Processing Model.....	50
Figure 3.1:	Participant Location and Frequency Map.....	101
Figure 4.1:	Expert Tag Counts by Format.....	126
Figure 4.2:	Novice Tag Counts by Format.....	127
Figure 4.3:	Box-plot Analysis of Number of Tags Generated by Participants.....	129
Figure 4.4:	Tag Cloud of all Replication of Metadata Tags.....	132
Figure 4.5:	Tag Cloud of all Format-Focused Tags.....	133
Figure 4.6:	Tag Cloud of all Subject-General Tags.....	134
Figure 4.7:	Tag Cloud of all Subject-Specific Tags.....	136
Figure 4.8:	Tag Cloud of all Content-Summary Tags.....	136
Figure 4.9:	Tag Cloud of all Context Tags.....	137
Figure 4.10:	Criticism Mail Letter 2, Wisconsin Historical Society, WHi-111271....	139
Figure 4.11:	Photograph 11, Wisconsin Historical Society, WHi-53596.....	140
Figure 4.12:	Tag Cloud of all Emotion Tags.....	140
Figure 4.13:	Tag Cloud of all Incorrect Tags.....	141
Figure 4.14:	Comparison Expert & Novice Tag Categories Percentage by Format...	144
Figure 4.15:	All Expert and Novice Tags by Category.....	146
Figure 4.16:	Expert and Novice Photograph Tags by Category.....	149
Figure 4.17:	Expert and Novice Document Tags by Category.....	151
Figure 4.18:	Proportions of Matching/Non-Matching of Tags to Unselected Metadata.....	166
Figure 4.19:	Tags and Unselected Metadata Matching User Query Terms Venn Diagram.....	169

Figure 6.1:	Research Findings and Implications.....	189
-------------	---	-----

LIST OF TABLES

Table 3.1:	Research Questions and Associated Data and Analysis.....	90
Table 3.2:	Sample Collection Minimal Metadata.....	95
Table 3.3:	Gender and Racial Characteristics of Participants.....	100
Table 3.4:	Religious Affiliation of Participants.....	102
Table 3.5:	Participants' Highest Completed Education Level.....	103
Table 3.6:	Average VAS Scores from Pre-Questionnaire.....	105
Table 3.7:	Coding Scheme Categories & Definitions.....	111
Table 4.1:	Aggregate Tag Counts by Users and Format.....	128
Table 4.2:	Examples of Photograph Tags by Category.....	130
Table 4.3:	Examples of Document Tags by Category.....	131
Table 4.4:	Most Frequent Replication of Metadata Tags.....	132
Table 4.5:	Most Frequent Format-Focused Tags.....	134
Table 4.6:	Most Frequent Subject—General Tags.....	135
Table 4.7:	Most Frequent Subject-Specific Tags.....	135
Table 4.8:	Most Frequent Content-Summary Tags.....	137
Table 4.9:	Most Frequent Context Tags.....	138
Table 4.10:	Most Frequent Emotion Tags.....	141
Table 4.11:	Most Frequent Incorrect Tags.....	141
Table 4.12:	Tag Counts and Percentages by Category and Format.....	143
Table 4.13:	Number and Percent of All Expert and Novice Tags by Category.....	145
Table 4.14:	Number and Percent of Expert and Novice Photograph Tags by Category.....	148

Table 4.15:	Number and Percent of Expert and Novice Document Tags by Category.....	150
Table 4.16:	Mann-Whitney U Test Comparison of Tagging Experience Statements.....	154
Table 4.17:	Mann-Whitney U Test Comparison of Tagging Consideration Statements.....	156
Table 4.18:	Existing Metadata Template for Groppi Papers.....	159
Table 4.19:	Number of Unselected Metadata Terms by Dublin Core Element.....	160
Table 4.20:	Number and Percent of Unselected Metadata Terms Matching User-Generated Tags by Dublin Core Element.....	161
Table 4.21:	Percent of Tags Matching Unselected Metadata Unique by Dublin Core Element.....	162
Table 4.22:	Proportion of Tags Matching Unselected Metadata.....	165
Table 4.23:	Comparison of All Collection Query Terms and Tags.....	168
Table 4.24:	Comparison of March on Milwaukee Query Terms and Tags.....	168

ACKNOWLEDGEMENTS

I must begin by expressing my deepest gratitude for the direction of my advisor and mentor, Dr. Iris Xie, whose direction and encouragement was central to my successful completion of this dissertation. I would also like to thank my dissertation committee members for volunteering numerous hours throughout the dissertation process: Dr. Paul Conway, Dr. Thomas Force, Dr. Wooseob Jeong, and Dr. Margaret E.I. Kipp. I will forever be grateful for their advice and comments. I would be remised to not make mention of several additional mentors who have assisted me during the doctoral journey including Dr. Kimberly Anderson, Dr. Carolyn Hank, and Dr. Aims McGuinness.

The dissertation research was partially funded through a Doctoral Research Award Grant Opportunity (DRAGO) from the School of Information Studies at the University of Wisconsin-Milwaukee. Without this funding, I would not have been able to incentivize participation in my study. Additionally, I need to acknowledge the Wisconsin Historical Society for providing reproduction permission for all of the images used in the dissertation. Along the same lines, several practitioners assisted my research during various stages including the staff of the University of Wisconsin-Milwaukee Archives. Most importantly was the assistance of Ann Hanlon, Head of Digital Collections and Initiatives at the UWM Libraries.

Finally, I would like to thank all of my friends and family for supporting me throughout this long endeavor. It was only with encouragement and a strong network that I could even dream of completing such as task. It will forever sadden me that my father, Edward Benoit, Jr. passed on before seeing me complete the dissertation; however, I know he

would be proud of the accomplishment. I would especially like to thank both Dr. Diana Belscamper and Sarah Ramdeen for being particularly helpful during the design of the study and providing excellent feedback. Although I could continue writing the names of all the people that provided some feedback, encouragement, or support, the list would continue for numerous pages; know that I recognize all of your roles in my success.

CHAPTER ONE: INTRODUCTION

The Internet revolution of the past two decades altered the information landscape, and how people interact with information in their daily lives. No longer were people restricted to using human intermediaries or gatekeepers with limited operation hours; rather, users could fill their information needs around the clock and through relatively simple information portals. Although the early years of the Internet offered significant improvements over traditional information-gathering behaviors, the static nature of Web 1.0 maintained some of the previous limitations of information access. The emergence of Web 2.0 gave a dynamic, interactive space where users collaborate, customize their information space, and engage with traditional information providers thereby creating a new information paradigm.

One of the more exciting aspects of the Web 2.0 movement is the growing popularity of crowdsourcing, or leveraging the wisdom of the crowd, to solve complex problems. Developed from the open-source movement, software developers and scientists initially used crowdsourcing for commercial projects such as creating more efficient recommendation algorithms for Netflix and citizen scientist projects such as Galaxy Zoo.¹ Crowdsourcing evolved to include user-generated indexing and social tagging, allowing users to arrange, rearrange, and access information through more personal methods while providing additional access points for other users, and what Weinberger

¹ David Weinberger, *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room* (New York: Basic Books, 2011).

calls the third order of order.² The inclusion of user participation within the creation and organization of knowledge alters the perception of professional knowledge and authority while offering an engagement with users addressing their personal needs.³

The archival community faced a massive backlog problem during the past twenty years, to the extent that some archives housed more unprocessed and, therefore, inaccessible, collections than processed ones. In response, Greene and Meissner proposed a drastic shift in both archival theory and practice toward the concept of “More Product, Less Process” or MPLP, and minimal processing.⁴ Briefly, MPLP strives toward identifying and implementing a minimal standard level of processing across collections thereby simultaneously decreasing the time required for processing while increasing the number of collections available to users. Minimal processing practice expanded throughout archival practice, from its origins with arrangement and description to digital archives, resulting in an increase of available collections both physically and digitally.

As one problem is solved, many more can be created. David Bearman and Margaret Hedstrom noted early in the study of electronic records, “In a period of down-sizing, right-sizing and just plain cutting back, the impact of new information

² David Weinberger, *Everything Is Miscellaneous: The Power of the New Digital Disorder* (New York: Times Books, 2007).

³ Charles Leadbeater and Debbie Powell, *We-Think: Mass Innovation, Not Mass Production* (London: Profile Books, 2008); Michael Lewis, *Next: The Future Just Happened* (New York: W. W. Norton & Company, 2001); Mirko Tobias Schäfer, *Bastard Culture! User Participation and the Extension of Cultural Industries* (Amsterdam: Amsterdam University Press, 2011); Clay Shirky, *Cognitive Surplus: Creativity and Generosity in a Connected Age* (New York: Penguin Press, 2010); Weinberger, *Everything Is Miscellaneous*; Weinberger, *Too Big to Know*.

⁴ Mark A. Greene and Dennis Meissner, “More Product, Less Process: Revamping Traditional Archival Processing,” *American Archivist* 68, no. 2 (2005): 208–263.

technologies is not the only challenge that archivists must confront.”⁵ The minimal processing technique in digital archives prioritizes the collection as a whole over individual items, specifically regarding metadata. The online collections provide only minimal metadata, typically at the series or folder level. The MPLP approach deviates from contemporary practice that describes digital archival materials at the item or record level. For example, each letter in a traditionally processed folder of digitized correspondence includes individualized descriptive metadata; the MPLP version of the same collection would only describe the folder as an aggregate with individual letters sharing duplicate metadata. While this replicates the experience of researchers in the physical archives, studies demonstrate an increasing demand for more description and access points from online users.

Reaching out to the same users for assistance and requesting them to help supplement minimally processed digital archives’ metadata through creation of tags could address this issue. Social tagging without some measure of control could, however, generate too many useless terms, thereby hindering access rather than increasing it. Additionally, archival users previously stated a preference for user-generated content-control mechanisms. While some suggest digital librarians and archivists simply approve/disapprove each tag, such a system requires too much oversight.⁶ I propose categorizing the users rather than the tags; specifically, permitting users who are subject-area experts (hereafter referred to as expert users) to tag the collections. I theorize that

⁵ David Bearman and Margaret Hedstrom, “Commentary Reinventing Archives for Electronic Records: Alternative Service Delivery Options,” in *Electronic Records Management Program Strategies*, ed. Margaret Hedstrom (Pittsburgh, PA: Archives & Museum Informatics, 1993), 82.

⁶ Edward Benoit III, “Digital Librarians’ Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations,” 2012, Manuscript in Preparation;

expert users provide more reliable tags, meeting the needs of institutions and improving access to the collections.

The digital archives adoption of minimal processing, or MPLP, returned archives to traditional description levels within digital archives with some issues. The introduction will continue by discussing the adoption process, the issues raised, and the proposed solutions. The remainder of this chapter describes the dissertation research problem, questions, hypothesis, and overall significance of the project.

1.1 Adaptions of Minimal Processing to Digital Archives

At the turn of the century, Cook argued that archival theory and practice underwent a significant paradigm shift in dealing with a postmodern world.⁷

Referencing Kuhn's ideas, Cook states:

[Kuhn] argued that radical changes occur in the interpretive framework for any scientific theory, which he called a paradigm shift, when answers to the research questions no longer explain sufficiently the phenomena being observed (in the archival case, recorded information and its creators) or when the practical methodologies based on theory from such observations no longer work (as they certainly do not for many archival activities, and not only coping with electronic records). The question and research focus, therefore, may remain "traditional in a paradigm shift;" the answers do not. And so it is with archives.⁸

⁷ Terry Cook, "Archival Science and Postmodernism: New Formulations for Old Concepts," *Archival Science* 1, no. 1 (2001): 3–24.

⁸ *Ibid.*, 5.

The emergence of minimal processing as a new solution to traditional problems of access and preservation falls within the archival postmodern paradigm shift.

As Greene notes, MPLP and minimal processing methods are not merely relevant for arrangement and description processes, but are applicable throughout archival practice.⁹ In his expanded discussion of MPLP, Greene disputes arguments that both born-digital and digitized records require item-level description within their associated metadata stating, “Why, in practice, should appraisal and description of electronic records be—or need to be—any different from that applied to analog material?”¹⁰ Furthermore, the backlog of electronic records significantly concerned Johnson since they “are far more fragile than their paper-based counterparts, and leaving them un-processed while an archivist creates a long and eloquent description endangers the record.”¹¹

Since users expect and demand more archival records to be digitally accessible, archivists must increase the number of digitized records by “abjuring item-level metadata” and archivists’ “fascination with individual documents.”¹² In rejecting item-level metadata, archivists and institutions reduce costs associated with digital archives creation, which in turn allows the digitization of additional collections. As one practitioner notes, “Every dollar spent to make [online] collections perfect is a dollar we’re not spending to get another collection online and to a larger potential audience.”¹³

⁹ Mark A. Greene, “MPLP: It’s Not Just for Processing Anymore,” *American Archivist* 73, no. 1 (2010): 175–203.

¹⁰ *Ibid.*, 192.

¹¹ Gregory P. Johnson, “Quality or Quantity: Can Archivists Apply Minimal Processing to Electronic Records?” 2007, 30, <http://www.ils.unc.edu/MSpapers/3267.pdf>.

¹² Mark A. Greene, “Doing Less Before It’s Done Unto You: Reshaping Workflows for Efficiency Before the Wolf Is at the Door,” *RBM: A Journal of Rare Books, Manuscripts and Cultural Heritage* 12, no. 2 (2011): 101.

¹³ Joshua Ranger, “More Bytes, Less Bite: Cutting Corners in Digitization,” 2008, <http://www.archivists.org/conference/sanfrancisco2008/docs/session701-ranger.pdf>.

Ranger further highlights that, “To cut costs in metadata, they cataloged items at the folder level instead of providing item-level metadata, giving the researcher enough data to locate a larger group of items they would be interested in.”¹⁴

A minimally processed digital archive, therefore, identifies the “golden minimum” metadata required to provide user access to the archival material. This level remains flexible for an entire repository, and may move from a series to subseries to folder level between collections depending on the collection. For example, folder-level metadata may be more suitable for a correspondence series containing several boxes and dozens of folders of correspondence; whereas limiting metadata at the series level for a correspondence series containing three folders would still provide adequate access to the digitized records. Following these procedures replicates contemporary archival methods for analog records and thereby allows users a similar experience to physically visiting the archives.

Assuming repositories would apply labor savings from a minimal processing approach towards increasing the number of digitized collections, the MPLP model provides a workable solution for the stagnated and shrinking budgets of modern archives. Additionally, the newly digitized materials may be accessed and used remotely, thereby addressing the rising demands of the 21st-century patron. By itself, however, digital archivists’ adoption of minimal processing does not take full advantage of content management systems such as OCLC’s CONTENTdm, since it mitigates the benefits of increased access points provided through record-level metadata.

¹⁴ Matt Gorzalski, “Minimal Processing: Its Context and Influence in the Archival Community,” *Journal of Archival Organization* 6, no. 3 (2008): 196.

Interestingly, Bearman and Hedstrom recognized the possibilities of minimal processing and electronic records early, stating:

In electronic records systems, metadata about the records and the configuration of permissions, views, and functions is created and controlled in the active data environment. In principle, this metadata if correctly specified could fully describe and document the records without post-hoc activity by the archivist.¹⁵

The abandonment of item-level description might better reflect the traditional approaches to description. Benson discusses the nature of early online systems of archival photographs, stating, “Item-level records for the majority of archival photographic materials were not common in early card catalog systems, so consequently there were no item-level records being migrated into first-generation online catalog systems.”¹⁶ Several researchers echo the MPLP approach without explicit mention. Deridder, Presnell and Walker, for example, sees “human-created item-level metadata,” as holding back the number of digitized materials.¹⁷ An OCLC report similarly states:

Vast quantities of digitized primary materials will trump a few superbly crafted special collections. Minimal description will not restrict use as much as limiting access to those who can show up in person. We must stop our slavish devotion to detail; the perfect has become the enemy of the possible.¹⁸

¹⁵ Bearman and Hedstrom, “Commentary Reinventing Archives for Electronic Records: Alternative Service Delivery Options,” 87.

¹⁶ Allen C. Benson, “The Archival Photograph and Its Meaning: Formalisms for Modeling Images,” *Journal of Archival Organization* 7, no. 4 (2009): 169.

¹⁷ Jody DeRidder, Amanda Presnell, and Kevin Walker, “Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis,” *American Archivist* 75, no. 1 (2012): 144.

¹⁸ Ricky Erway and Jennifer Schaffner, *Shifting Gears: Gearing up to Get Into the Flow* (Dublin, OH: OCLC Programs and Research, 2007), <http://www.oclc.org/programs/publications/reports/2007-02.pdf>.

Although the MPLP approach to digital archives presents digital surrogates of archival materials in a similar fashion to their use in physical archives, many users may have difficulties navigating the collection (specifically, users without archival research experience). Altman and Nemmers have found that users prefer item-level descriptions and have difficulty following online finding aids (Prom provided similar results).¹⁹ As Deridder, Presnell, and Walker reflected on their decision to abandon item-level description, they state, “A drawback, however, is that this method of Web delivery may currently be more suitable for scholars than for students.”²⁰ Furthermore, when looking at the use of archival resources, Ham et al. suggests, “Other user groups may frame questions different from those of historians.”²¹

The minimally processed digital archives could frustrate non-traditional archival users who approach digital archives similarly to other Web-based information retrieval systems. According to Xie, most users “are only willing to devote a small amount of time to evaluate [search] results.”²² Additionally, she states, “In digital environments, interaction with results has become a major component of information retrieval interaction. Users interact with results to find information to solve their problems; these results lead them to search for needed information or to find new ideas to reformulate their queries if the results fail to provide relevant information.”²³ In comparing search result list and document evaluation, Xie and Benoit recommend additional evaluation

¹⁹ Burt Altman and John R. Nemmers, “The Usability of On-Line Archival Resources: The Polaris Project Finding Aid,” *American Archivist* 64, no. 1 (2001): 121–131; Christopher J. Prom, “User Interactions with Electronic Finding Aids in a Controlled Setting,” *American Archivist* 67, no. 2 (2004): 234–268.

²⁰ DeRidder, Presnell, and Walker, “Leveraging Encoded Archival Description for Access to Digital Content.”

²¹ F. Gerald Ham et al., “Is the Past Still Prologue?: History and Archival Education,” *American Archivist* 56, no. 4 (October 1, 1993): 718–729.

²² Iris Xie, *Interactive Information Retrieval in Digital Environments* (Hershey, PA: IGI Pub., 2008).

²³ *Ibid.*, xiv.

information presented with search results for adequate evaluation.²⁴ With only minimal metadata to guide their evaluations, however, users may either accidentally pass over relevant documents or slowly evaluate each record regardless of metadata descriptions.

Social tagging within digital collections has gained interest in the past decade.²⁵

The inclusion of tags within digital archives could reintroduce some of the access points lost when utilizing a minimal processing approach. Previous studies of Web 2.0 tools within online archival offerings (both collections and finding aids) suggest both users and archivists remain reluctant to leverage unmitigated crowdsourcing.²⁶ Users distrust the tags generated from other general users; however, they would consider using information created by so-called expert researchers and users.²⁷ I posit the best solution for

²⁴ Iris Xie and Edward Benoit III, "Search Result List Evaluation versus Document Evaluation: Similarities and Differences," *Journal of Documentation* 69, no. 1 (2013): 49–80.

²⁵ Maristella Agosti et al., "Annotation As a Support to User Interaction for Content Enhancement in Digital Libraries," in *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '06* (New York, NY, USA: ACM, 2006), 151–154, <http://doi.acm.org/10.1145/1133265.1133296>; David Bearman and Jennifer Trant, "Social Terminology Enhancement through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections," *D-Lib Magazine* 11, no. 9 (2005), <http://www.dlib.org/dlib/september05/bearman/09bearman.html>; Krystyna K. Matusiak, "Towards User-Centered Indexing in Digital Image Collections," *OCLC Systems & Services: International Digital Library Perspectives* 22, no. 4 (2006): 283–298; Michelle Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project* (The Library of Congress, 2008), http://www.loc.gov/tr/print/flickr_report_final.pdf; Jennifer Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept," in *Art Museums: Proof of Concept. New Review of Hypermedia and Multimedia* 12, no. 1 (2006): 83–105; Helena Zinkham and Michelle Springer, "Taking Photographs to the People: The Flickr Commons Project and the Library of Congress," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 102–115.

²⁶ Scott R. Anderson and Robert B. Allen, "Envisioning the Archival Commons," *American Archivist* 72, no. 2 (2009): 383–400; Adam Crymble, "An Analysis of Twitter and Facebook Use by the Archival Community," *Archivaria* 70 (2010): 125–151; Magia Ghetu Krause and Elizabeth Yakel, "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *American Archivist* 70, no. 2 (2007): 282–314; Frank Upward, Sue McKemmish, and Barbara Reed, "Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures," *Archivaria* 72 (2011): 197–237; Elizabeth Yakel, "Inviting the User into the Virtual Archives," *OCLC Systems & Services* 22, no. 3 (2006): 159–163.

²⁷ Jodi Allison-Bunnell, Elizabeth Yakel, and Janet Hauck, "Researchers at Work: Assessing Needs for Content and Presentation of Archival Materials," *Journal of Archival Organization* 9, no. 2 (2011): 67–104.

maintaining the item levels of access within a minimally processed digital archive is the inclusion of tags created by domain experts.

The inclusion of tags meets the needs of a diverse user base. Bearman and Hedstrom consider the potential for community involvement of “users in problem solving and service delivery within a clearly articulated framework of principles and standards...to achieve mutually desired ends.”²⁸ Additionally, this framework will help archives deal with the inherent problems of description:

Classification systems, thesauri, and other metadata encoding schemes developed within one worldview do not include the concepts and terms needed to classify and name entities within another. Metadata standards built within continuum frameworks have been designed to support an enduring view of records and their contexts, capturing the dynamic and changing relationships between the multiple entities in the recordkeeping and archiving landscape.²⁹

1.2 Research Problem, Questions, and Hypotheses

The high costs of creating and maintaining digital archives precluded many archives from providing users with digital content or increasing the amount of digitized materials. Studies have shown users increasingly demand immediate online access to archival materials with detailed descriptions (access points). The adoption of minimal processing to digital archives limits the access points at the folder or series level rather than the item-level description users’ desire. User-generated content such as tags, could supplement the minimally processed metadata, though users are reluctant to trust or use

²⁸ Bearman and Hedstrom, “Commentary Reinventing Archives for Electronic Records: Alternative Service Delivery Options,” 91.

²⁹ Upward, McKemmish, and Reed, “Archivists and Changing Social and Information Spaces,” 230.

unmediated tags. This dissertation project explores the potential for controlling/mediating the supplemental metadata from user-generated tags through inclusion of only expert domain user-generated tags. Furthermore, the dissertation investigates the following research questions and associated hypotheses:

- Research Question 1(a): What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?
 - H₁: The number of tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₂: The number of photographic tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₃: The number of document tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₄: The proportion of tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₅: The proportion of photographic tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₆: The proportion of document tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
- Research Question 1(b): Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?

- H₇-H₉: Expert and novice users' opinions of the tagging experience are different for ease of tagging in general (H₇); difficulty in tagging documents compared to photographs (H₈); and difficulty in tagging photographs compared to documents (H₉).
- H₁₀-H₁₄: Expert and novice users' opinions of the considerations for the creation of tags are different for how others would find the item (H₁₀); how the tagger (user) would find the item (H₁₁); the content of the tagged item (H₁₂); the format of the tagged item (H₁₃); and other users' tags (H₁₄).
- Research Question 2 (a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?
- Research Question 2 (b): Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?
 - H₁₅: The proportion of tags matching unselected metadata is affected by the user's domain knowledge.
- Research Question 3 (a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?
- Research Question 3 (b): Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?
 - H₁₆: The proportion of tag terms matching users' query log terms is affected by user's domain knowledge.

1.3 Research Design

The dissertation project addresses the research questions and hypotheses through a mixed-methods, quasi-experimental design focused on tag generation within a sample minimally processed digital archive. The study used a sample collection of fifteen documents and fifteen photographs from the Groppi Papers portion of the existing *The March on Milwaukee Civil Rights History Project* (hereafter called *March on Milwaukee*) at the Digital Collections at the University of Wisconsin-Milwaukee Libraries. The fifteen documents were equally divided between three subgroupings (hate mail, support mail, and criticism mail). The sample collection selected and extracted a shared set of minimally processed metadata from the existing *March on Milwaukee* collection.

Sixty participants divided into two groups (novices and experts) based on assessed prior knowledge of the Civil Rights movement in Milwaukee generated tags for fifteen documents and fifteen photographs (a minimum of one tag per object). Participants completed a pre-questionnaire identifying prior knowledge, and use of social tagging and archives. Additionally, participants provided their opinions regarding factors associated with tagging including the tagging experience and considerations while creating tags through structured and open-ended questions in a post-questionnaire.

An open-coding analysis of the created tags developed a coding scheme of six major categories and six subcategories. Application of the coding scheme categorized all generated tags. Additional descriptive statistics summarized the number of tags created by each domain group (expert, novice) for all objects and divided by format (photograph, document). T-tests and Chi-square tests explored the associations (and associative

strengths) between domain knowledge and the number of tags created or types of tags created for all objects and divided by format.

The subsequent analysis compared the tags with the metadata from the *March on Milwaukee* collection not displayed within the sample collection participants used. The comparison with this so-called unselected metadata explored the potential for duplicating traditional item-level description by including tags within a minimally processed digital archive. Descriptive statistics summarized the proportion of tags matching unselected metadata and Chi-square tests analyzed the findings for associations with domain knowledge.

Finally, the author extracted existing users' query terms from one month of the Digital Collections at the University of Wisconsin-Milwaukee Libraries' server-log data, thereby creating two lists of query terms. One list included searches across multiple collections (including the *March on Milwaukee*), and the second list included only searches of the *March on Milwaukee*. The generated-tags and unselected metadata were compared with both query term lists identifying the potential information retrieval possibilities of tagging within a minimally processed digital archive. Descriptive statistics summarized the proportion of tags and unselected metadata matching query terms, and Chi-square tests analyzed the findings for associations with domain knowledge.

1.4 Significance of Dissertation Project

Changing times and technology require innovative solutions. The dissertation project addresses the need for increasing the number of digital collections available while meeting users' need for item-level access points to access the digital collections.

Additionally, the project explores the difference between archival description and user driven folksonomies. This project falls into what Conway laid out as stage 5 in his framework for studying archival users, the need for experimental research of innovative tools and solutions to contemporary issues.³⁰ Similarly, the dissertation project answers Speck's call that, "More studies should be done to ascertain the benefits of using social interaction tools for improving both finding aids and the overall online presence of archives."³¹ Furthermore, considering participatory archives, including social tagging, Flinn called for continued research "to find out what works and what does not, to explore how the reliability of the entries is to be gauged, to examine the continued role for professional mediation, and what...the relationship to the professional catalogue is."³²

While minimal processing has previously been adapted for digital archives, it did not address the ongoing calls for increased access points and more description of archival materials. In fact, through leveraging the MPLP practice, archives reduce the access points and description associated with digital archives (as well as other aspects). The dissertation project addresses these deficiencies through empirical testing of a potential crowdsourcing solution. The mixed-methods, quasi-experimental-designed study controls variables and provides a reliable basis for exploring expert-user-generated tags.

The findings of the dissertation project have significant implications for archival theory. The project enhances the understanding of the minimal processing model's ongoing role in the shifting landscape of archives in the digital era. The results reinforce

³⁰ Paul Conway, "Facts and Frameworks: An Approach to Studying the Users of Archives," *American Archivist* 49, no. 4 (1986): 393–407.

³¹ Jason G. Speck, "Protecting Public Trust: An Archival Wake-Up Call," *Journal of Archival Organization* 8, no. 1 (2010): 31–53.

³² Andrew Flinn, "An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge," *Ariadne* 62 (2010), <http://www.ariadne.ac.uk/issue62/flinn>.

or broaden the findings of previous archival studies, specifically those focused on participatory user engagement and calls for additional research into user-created content. The findings further illuminate how users interpret archival materials through the analysis of the tags both novices and experts create to describe materials. Most importantly, the dissertation adds to the ongoing postmodern movement with heterogeneous description from user-generated tags. Finally, the dissertation's results also provide theoretical implications based on previous research into social tagging by both reinforcing and disputing prior findings.

The practical implications of the findings add to the dissertation's significance through providing concrete recommendations for future use of social tags within minimally processed digital archives. Specifically, the association between types of tags and prior domain knowledge requires that repositories alter the user tagging requirements based on the archive's desired use of the tags. For example, if an archive prefers content-summary tags, it should consider restricting tagging to experts. Additional findings negate concern over incorrect tags, while reinforcing issues of tags replicating already existing metadata. The dissertation study's participants provide suggestions for future system development and recommendations regarding motivating tag creation. Finally, the comparison of generated tags with unselected metadata and query terms implies tagging alone cannot replace item-level description, and documents would benefit from additional content-driven metadata.

1.5 Summary of Dissertation

The following chapters will further outline the dissertation project and discuss its results. Chapter 2 examines the interpretation of digital archives and contextualizes the

dissertation's position in archival theory and literature. This chapter also outlines the development of the minimal processing model framework of the dissertation, discusses the existing social tagging literature, and the gaps or limitations therein. Chapter 3 discusses and justifies the particular methodology employed for the dissertation project. Chapter 4 presents the study's findings and Chapter 5 discusses the implications of those findings. Finally, Chapter 6 summarizes the finding of the dissertation, and highlights future research directions.

CHAPTER TWO: LITERATURE REVIEW

The development of archival theory and practice toward digital archives during the past decades highlights the fragility and ephemeral nature of electronic records;³³ raises significant issues about how to best represent archival records in order to provide access;³⁴ and has resulted in major shifts in discussions on what to collect and what defines a record in the digital age.³⁵ Early research focused on the state of electronic, or born-digital, records in archives, with an emphasis on the unpreparedness of archives to handle born-digital records.³⁶ While many recognized the promises of new methods for

³³ David Bearman, "Reality and Chimeras in the Preservation of Electronic Records," *D-Lib Magazine* 5, no. 4 (1999): 1–5; Jean Dryden, "The Open Archival Information System Reference Model," *Journal of Archival Organization* 7, no. 4 (October 2009): 214–217; Lilly Koltun, "The Promise and Threat of Digital Options in an Archival Age," *Archivaria* 47, no. 1 (1999): 114–135.

³⁴ Eric Ketelaar, "Commentary on 'Archival Strategies': The Archival Image," *American Archivist* 58, no. 4 (October 1, 1995): 454–456; Sarah Tyacke, "Archives in a Wider World: The Culture and Politics of Archives," *Archivaria* 52, no. 1 (2001): 1–25; Jane Zhang, "The Principle of Original Order & the Organization and Representation of Digital Archives" (Ph.D., Simmons College, 2010), <http://search.proquest.com.ezproxy.library.yorku.ca/pqdtft/docview/861937648/abstract/CA92C32440F34649PQ/1?accountid=15182>.

³⁵ Richard J. Cox, "Yours Ever (well, Maybe): Studies and Signposts in Letter Writing," *Archival Science* 10, no. 4 (2010): 373–388; Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES," *Archival Science* 6, no. 1 (2006): 13–68; Berndt Fredriksson, "Postmodernistic Archival Science — Rethinking the Methodology of a Science," *Archival Science* 3, no. 2 (2003): 177–197; Margaret Hedstrom, "Archives, Memory, and Interfaces with the Past," *Archival Science* 2, no. 1–2 (2002): 21–43; Christopher A Lee, ed., *I, Digital: Personal Collections in the Digital Era* (Chicago: Society of American Archivists, 2011); Tom Nesmith, "Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives," *American Archivist* 65, no. 1 (2002): 24–41.

³⁶ Bearman and Hedstrom, "Commentary Reinventing Archives for Electronic Records: Alternative Service Delivery Options"; David Bearman, "New Models for Management of Electronic Records," in *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, ed. David Bearman, vol. 2 (Pittsburgh, PA: Archives & Museum Informatics, 1994), 278–292, http://www.archimuse.com/publishing/electronic_evidence/ElectronicEvidence.Ch10.pdf; David Bearman, "Archival Principles and the Electronic Office," in *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, ed. David Bearman (Pittsburgh, PA: Archives & Museum Informatics, 1994), 145–176, http://www.archimuse.com/publishing/electronic_evidence/ElectronicEvidence.Ch10.pdf; David Bearman, "Information Technology Standards and Archives," in *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, ed. David Bearman (Pittsburgh, PA: Archives & Museum Informatics, 1994), 210–221, http://www.archimuse.com/publishing/electronic_evidence/ElectronicEvidence.Ch10.pdf; Susan E. Davis, "Electronic Records Planning in 'Collecting' Repositories," *American Archivist* 71, no. 1 (2008): 167–189; Patricia Sleeman, "Notes and Communications It's Public Knowledge: The National Digital Archive of Datasets," *Archivaria* 58, no. 1 (2004): 173–200; Lisl Zach and Marcia Frank Peri, "Practices for College

access, aggregating, and analyzing records,³⁷ others were concerned over the issues of authenticating electronic records,³⁸ and preservation.³⁹ Along with the changing nature of records, digital archives brought questions of the limitations of the traditional life-cycle-of-records approach, and a recommended move towards a continuum model.⁴⁰

As archives moved toward online digital archives, the research shifted toward exploring the difference from physical archives;⁴¹ the altered relationships between archivist and user/researcher;⁴² and questions of provenance.⁴³ Some researchers summarized the state of digital archives,⁴⁴ digitization problems encountered, and

and University Electronic Records Management (ERM) Programs: Then and Now,” *American Archivist* 73, no. 1 (2010): 105–128.

³⁷ Bearman and Trant, “Social Terminology Enhancement through Vernacular Engagement”; Christopher French, “Computerizing London’s Eighteenth-Century Maritime Activity,” *Archives: Journal of BRA* 22, no. 97 (1997): 130–140.

³⁸ Luciana Duranti and Randy Preston, eds., *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008), http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_complete.pdf; Anne J. Gilliland-Swetland, “Testing Our Truths: Delineating the Parameters of the Authentic Archival Electronic Record,” *American Archivist* 65, no. 2 (2002): 196–215; InterPARES Project, *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, 2012, <http://www.interpares.org/book/index.cfm>; InterPARES Project, *InterPARES 3 Project*, 2012, http://www.interpares.org/ip3/ip3_index.cfm.

³⁹ Terry D. Baxter, “Going to See the Elephant: Archives, Diversity, and the Social Web,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 274–303; Paul Conway, *Preservation in the Digital World* (Washington, DC: Commission on Preservation and Access, 1996), <http://www.clir.org/pubs/reports/conway2/>.

⁴⁰ Sue McKemmish and Michael Piggott, eds., *The Records Continuum: Ian Maclean and Australian Archives: First Fifty Years* (Clayton, Australia: Ancora Press, 1994); Sue McKemmish, “Placing Records Continuum Theory and Practice,” *Archival Science* 1, no. 4 (2001): 333–359.

⁴¹ Dayna Holz, “Technologically Enhanced Archival Collections: Using the Buddy System,” *Journal of Archival Organization* 4, no. 1–2 (2007): 29–44; Cory Nimer and J. Gordon Daines, “What Do You Mean It Doesn’t Make Sense? Redesigning Finding Aids from the User’s Perspective,” *Journal of Archival Organization* 6, no. 4 (2008): 216–232.

⁴² Charles W. J. Withers and Andrew Grout, “Authority in Space?: Creating a Digital Web-Based Map Archive,” *Archivaria* 61, no. 61 (2006), <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12533>.

⁴³ Margaret Hedstrom, “Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible,” *Archivaria* 1, no. 36 (1993): 53–63.

⁴⁴ Karen Gracy, “Distribution and Consumption Patterns of Archival Moving Images in Online Environments,” *American Archivist* 75, no. 2 (2012): 422–455; Mary Samouelian, “Embracing Web 2.0: Archives and the Newest Generation of Web Applications,” *American Archivist* 72, no. 1 (2009): 42–71.

solutions.⁴⁵ Recently, several researchers focused on user studies looking at how archival use changes when shifted from physical to digital;⁴⁶ the use of electronic finding aids;⁴⁷ the lack of discoverability of digital archives;⁴⁸ and general calls for more users studies to inform digitization efforts.⁴⁹

Along with digital archives research and electronic records, archivists and researchers began exploring the potential uses of Web 2.0 tools within the framework of Archives 2.0. This research highlights new interactions between user and archivist/archives with Web 2.0 tools.⁵⁰ The majority of literature discusses an increased role for users with case studies exploring the potential of user-generated content and flexibility;⁵¹ GIS mapping;⁵² digital repatriation;⁵³ and capturing user knowledge and

⁴⁵ Marjan Balkestein and Heiko Tjalsma, "The ADA Approach: Retro-Archiving Data in an Academic Environment," *Archival Science* 7, no. 1 (2007): 89–105; Devin Becker and Collier Nogue, "Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers," *American Archivist* 75, no. 2 (2012): 482–513; Laura Carroll et al., "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72, no. Fall (2011): 61–92; Andrea Watson and P. Toby Graham, "CSS Alabama 'Digital Collection': A Special Collections Digitization Project," *American Archivist* 61, no. 1 (1998): 124–134.

⁴⁶ Wendy M. Duff et al., "Archivists' Views of User-Based Evaluation: Benefits, Barriers, and Requirements," *American Archivist* 71, no. 1 (2008): 144–166; Wendy M. Duff and Joan M. Cherry, "Archival Orientation for Undergraduate Students: An Exploratory Study of Impact," *American Archivist* 71, no. 2 (2008): 499–529; Elizabeth Shepard, "Digitizing a Photographic Collection in a Midsize Repository: A Case Study," *Journal of Archival Organization* 2, no. 4 (2004): 67–82.

⁴⁷ Altman and Nemmers, "The Usability of On-Line Archival Resources"; Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting."

⁴⁸ Allison-Bunnell, Yakel, and Hauck, "Researchers at Work."

⁴⁹ Duff et al., "Archivists' Views of User-Based Evaluation"; Anne J. Gilliland-Swetland, "An Exploration of K-12 User Needs for Digital Primary Source Materials," *American Archivist* 61, no. 1 (1998): 136–157.

⁵⁰ Anderson and Allen, "Envisioning the Archival Commons"; Crymble, "An Analysis of Twitter and Facebook Use by the Archival Community"; Kate Theimer, *A Different Kind of Web: New Connections between Archives and Our Users* (Chicago: Society of American Archivists, 2011); Kate Theimer, "What Is the Meaning of Archives 2.0?," *American Archivist* 74, no. 1 (2011): 58–68; Upward, McKemmish, and Reed, "Archivists and Changing Social and Information Spaces"; Yakel, "Inviting the User into the Virtual Archives."

⁵¹ Krause and Yakel, "Interaction in Virtual Archives."

⁵² Deborah Boyer, Robert Cheetham, and Mary Johnson, "Using GIS to Manage Philadelphia's Archival Photographs," *American Archivist* 74, no. 2 (2011): 652–663.

⁵³ Kimberly Christen, "Opening Archives: Respectful Repatriation," *American Archivist* 74, no. 1 (2011): 185–210.

encouraging user participation.⁵⁴ Although users had the potential for expanded roles, some research notes users are willing to use other users' generated content, but did not want to leave their own.⁵⁵

2.1 Defining Digital Archives

The relative infancy, and dynamic nature of born-digital and digitized records precludes a clear, concise, and universally agreed upon definition of digital archive. The potential defining characteristics range from an all-encompassing approach with the inclusion of born-digital and digitized materials (or any combination thereof) from both single and multiple archival collections to narrow approaches limiting digital archives, to born-digital materials from a single archival collection. The particular definition utilized by specific authors depends on the purpose and framework of their studies and analyses. The dissertation project is no exception and must therefore set its use of digital archives within a particular framework for meaningful discussion of the findings. The sample collection used during the quasi-experimental design must also fit within the definitional framework.

For the purpose of the dissertation project, therefore, a digital archive is defined and limited to curated online collections of digitized materials selected from a single or multiple existing physical archival collection(s), which adheres to the archival principles of provenance and original order, and is, at a minimum, arranged and described following contemporary best archival practices. This definition excludes collections of born-digital

⁵⁴ Gwen Evans and Susannah Cleveland, "Moody Blues: The Social Web, Tagging, and Nontextual Discovery Tools for Music," *Music Reference Services Quarterly* 11, no. 3–4 (2008): 177–201; Katie Shilton and Ramesh Srinivasan, "Participatory Appraisal and Arrangement for Multicultural Archival Collections," *Archivaria* 63, no. 1 (2007): 87–101.

⁵⁵ Allison-Bunnell, Yakel, and Hauck, "Researchers at Work."

materials, digitization of an entire analog collection, online finding aids, and online descriptions of archival materials without digital surrogates of the described objects. The definition includes selections from multiple repositories and multiple formats of objects (e.g., textual, image, audio, moving image). The sample digital archives used for the dissertation project (discussed in detailed within the methodology chapter) fulfills the specified characteristics since it contains digitized correspondence and photographs selected as representative of an existing physical collection, and maintains the physical collection's arrangement and description through aggregation into compound digital objects (similar to folder-level arrangement).

A significant challenge within archival literature is arriving at a consensus definition for terminology. Archivists and researchers continue to debate foundational principles, such as provenance and original order, and their positions within the archival framework after over a century of theory and practice. It is no surprise, therefore, that the relatively new idea of digital archives is not an exception to the rule. The following section examines the variety of approaches and definitions toward digital archives.

A growing concern is the adaptation of “archives” within technological terms, such as “archiving a file,” or the “archive” button in Gmail. As Tyacke notes, “Perhaps because the images and information are not in book form, the term “-archive-” seems to have become far more common in “-IT-speak-” than the term “-library-” which remains more solidly positioned with books, specific place, and information.”⁵⁶ Koltun states, “Digital players have begun to take over our word—‘archive’ (in the singular) has a

⁵⁶ Tyacke, “Archives in a Wider World.”

sudden, new cachet, as in the ‘digital archive,’ or the ‘archiving’ of ‘data.’”⁵⁷ Tension remains between the library, museum, and archival worlds over the delineations between the digital representations of each. Cunningham places the archival viewpoint simply stating, “Just as archives are different from libraries and museums, so, too, should digital archives be different from digital libraries and museums.”⁵⁸ Moreover, he states, “Digital archives are at risk of being managed just like vanilla digital libraries, thus dumbing down the peculiar challenges and complexities of preserving records.”⁵⁹

Simultaneously, the outlook from the museum side appears to highlight shared overall goals between digital libraries, archives and museums. In one of the earliest discussions of digital or virtual museums, and in discussing the first conference on hypermedia and interactivity in museums, Bearman states, “Since the early 20th century, museums have strived to be more than ‘cabinets of curiosities’ to be viewed passively.”⁶⁰ The ‘virtual museum’ “enable[s] explorations of the unique, the remote and the difficult to perceive, which can take place in a school, in the home, or on the street as easily as in the museum itself.”⁶¹ Projects such as the CSS Alabama Digital Collection combine both library and archival materials while respecting the divisions between them. This is reflected in the project’s collection-development policy stating that the priority “would be, first, unique and rare documents from manuscript holdings, images of the ship, its plans and personnel, relevant items from contemporaneous published sources; and,

⁵⁷ Koltun, “The Promise and Threat of Digital Options in an Archival Age,” 119.

⁵⁸ Adrian Cunningham, “Digital Curation/Digital Archiving: A View from the National Archives of Australia,” *American Archivist* 71, no. 2 (2008): 532.

⁵⁹ *Ibid.*, 533.

⁶⁰ David Bearman, “Interactive and Hypermedia in Museums,” in *Hypermedia & Interactivity in Museums, Proceedings of an International Conference* (Pittsburgh, PA: Archives & Museum Informatics, 1991), 1, http://www.archimuse.com/publishing/ichim_91.html.

⁶¹ *Ibid.*, 2.

second, as many copyright-free monographs as time and energy might permit.”⁶²

Additionally, Monks-Leeson sees the online archives as a method for bringing together both library and archival materials “which had otherwise been scattered across different libraries and archives, and thus both restore and establish contextual bonds that would have remained hidden.”⁶³

Despite these possible shared goals, several researchers warn against a merged outlook on digital materials. In his discussion of the different approaches to digital collections from the library and archives perspectives, Sterling Coleman summarizes the major issues as a “conflict over two fundamental questions that strike at the heart of collection development and collection management: What is a collection and how shall it be arranged?”⁶⁴ From a librarian perspective, a collection is comprised of topically arranged and gathered materials from multiple sources, whereas an archival perspective regards provenance and original order as primary concerns. Coleman warns against the librarian practice of integrating archival materials within a digital library without respecting archival principles. He suggests multiple solutions to this problem, including separating digital libraries and digital archives within a particular institution, although this may lead to “initial confusion that would come from a user who would have to cross between two different databases and interfaces.”⁶⁵ Another suggested approach focuses on selecting the majority visual archival collections for inclusion within digital libraries.

⁶² Watson and Graham, “CSS Alabama ‘Digital Collection,’” 127.

⁶³ Emily Monks-Leeson, “Archives on the Internet: Representing Contexts and Provenance from Repository to Website,” *American Archivist* 74, no. 1 (2011): 55.

⁶⁴ Sterling Coleman, “The Archival and Library Viewpoints of a Collection in a Digital Environment: Is There Any Room for Compromise?” *Journal of Archival Organization* 2, no. 1 (2004): 104–105.

⁶⁵ *Ibid.*, 107.

Focusing on defining characteristics of digital archives also involves separating out what digital archives are not; for example, Cunningham states, “Digital archiving cannot just be end-of-life-cycle collection management” and that, “Just as archival operations are more than preservation, digital archives are more than digital preservation.”⁶⁶ At the same time, “The OAIS Reference Model uses ‘digital archive’ to mean the organization responsible for digital preservation.”⁶⁷ Furthermore, the Research Library Group views both digital archives and preservation linked by defining digital preservation as, “the managed activities necessary for ensuring both the long-term maintenance of a bytestream and continued accessibility of its contents.”⁶⁸ Although, importantly, Jantz and Giarlo note, “The definition therefore does not apply to virtually all of the born-digital resources that have no corresponding physical representation.”⁶⁹ Authors such as Oliver, Chawner, and Liu use digital archives to mean only born-digital records, and not those digitized.⁷⁰

Digital archives and digital archiving are also not synonymous with digital curation. According to Cunningham, “The DCC defines the phrase as ‘maintaining and adding value to a trusted body of digital information for current and future use.’”⁷¹ He continues by further delineating the differences between digital archives and digital curation, stating:

⁶⁶ Cunningham, “Digital Curation/Digital Archiving,” 533, 540.

⁶⁷ Research Library Group, *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, CA: Research Libraries Group, 2002), 3, <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.

⁶⁸ Ibid.

⁶⁹ Ronald Jantz and Michael Giarlo, “Digital Archiving and Preservation: Technologies and Processes for a Trusted Repository,” *Journal of Archival Organization* 4, no. 1 (2006): 195.

⁷⁰ Gillian Oliver, Brenda Chawner, and Hai Ping Liu, “Implementing Digital Archives: Issues of Trust,” *Archival Science* 11, no. 3–4 (2011): 311–327.

⁷¹ Cunningham, “Digital Curation/Digital Archiving,” 531.

Just as archiving (the management of archives and records) is but one form of curation, so too is digital archiving just one form of digital curation. Yet the two terms are so often used interchangeably as to appear to be synonymous. They are not. Digital curation of archival materials is not just about digital collection management. In fact, the curation of digital records is a sufficiently distinct curatorial activity as to warrant the use of a different term—digital archiving.⁷²

Interestingly, Cunningham continues by placing digital librarianship also under the scope of digital curation by stating, “Included within the definition of digital curation are noble endeavors of digital preservation, digital librarianship, data management.”⁷³ The concept of digital curation, therefore, is a broader, overarching perspective of digital information. According to Yakel, “The active involvement of information professionals in the management, including the preservation, of digital data for future use.”⁷⁴ And finally, Lee and Tibbo state, “‘Digital curation’ is less wedded to specific institutional types than phrases such as ‘digital archives’ or ‘digital libraries.’”⁷⁵

After reviewing the literature, many different terms arise for a similar construct.

These include:

- *digital archives*;⁷⁶

⁷² Ibid.

⁷³ Ibid.

⁷⁴ Christopher A. Lee and Helen Tibbo, “Where’s the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills,” *Archivaria* 72 (2011): 124.

⁷⁵ Ibid., 126.

⁷⁶ Jean Dryden, “Measuring Trust: Standards for Trusted Digital Repositories,” *Journal of Archival Organization* 9, no. 2 (2011): 127–130; Duff et al., “Archivists’ Views of User-Based Evaluation”; Patricia Galloway, “Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice,” *Archivaria* 72 (2011): 169–196; Monks-Leeson, “Archives on the Internet,” –; Oliver, Chawner, and Liu, “Implementing Digital Archives”; Robert B. Townsend, “Old Divisions, New Opportunities: Historians and Other Users Working with and in Archives,” in *A Different Kind of Web*:

- *digital collection*;⁷⁷
- *digital exhibition*;⁷⁸
- *e-archives or electronic archives*;⁷⁹
- *online archives*;⁸⁰
- *online exhibition*;⁸¹
- *virtual archives*;⁸²
- *virtual collection*;⁸³
- *virtual exhibits*;⁸⁴ and
- *website archives*.⁸⁵

New Connections between Archives and Our Users, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 213–232; Withers and Grout, “Authority in Space?”; Jane Zhang, “Original Order in Digital Archives,” *Archivaria* 74, no. 1 (2012): 167–193.

⁷⁷ Samouelian, “Embracing Web 2.0”; Zhang, “Original Order in Digital Archives.”

⁷⁸ Samouelian, “Embracing Web 2.0.”

⁷⁹ Bearman, “New Models for Management of Electronic Records”; Jane Greenberg, “The Applicability of Natural Language Processing (NLP) to Archival Properties and Objectives,” *American Archivist* 61, no. 2 (1998): 400–425; Tyacke, “Archives in a Wider World.”

⁸⁰ Monks-Leeson, “Archives on the Internet.”

⁸¹ Joy Palmer and Jane Stevenson, “Something Worth Sitting for? Some Implications of Web 2.0 for Outreach,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 1–21.

⁸² Karen Anderson et al., “Teaching to Trust: How a Virtual Archives and Preservation Curriculum Laboratory Creates a Global Education Community?,” *Archival Science* 11, no. 3–4 (2011): 349–372; David Bearman, “Virtual Archives” (presented at the ICA Meeting, Beijing, China, 1996), <http://web.archive.org/web/19990427133904/http://www.lis.pitt.edu/~nhprc/prog6.html>; Robin L. Chandler, “Building Digital Collections at the OAC,” *Journal of Archival Organization* 1, no. 1 (2002): 93–103; Peter Doorn and Heiko Tjalsma, “Introduction: Archiving Research Data,” *Archival Science* 7, no. 1 (2007): 1–20; Andrea Medina-Smith, “Going Where the Users Are: The Jewish Women’s Archive and Its Use of Twitter,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 65–74; Mattie Taormina, “The Virtual Archives: Using Second Life to Facilitate Browsing and Archival Literacy,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 42–53; Withers and Grout, “Authority in Space?”; Yakel, “Inviting the User into the Virtual Archives.”

⁸³ Chandler, “Building Digital Collections at the OAC”; Davis, “Electronic Records Planning in ‘Collecting’ Repositories”; Bradley D. Westbrook, “Prospecting Virtual Collections,” *Journal of Archival Organization* 1, no. 1 (2002): 73–80.

⁸⁴ James Gerencser, “New Tools Equal New Opportunities: Using Social Media to Achieve Archival Management Goals,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 159–179.

⁸⁵ Monks-Leeson, “Archives on the Internet,” and Samouelian, “Embracing Web 2.0.”

Additionally, many authors alternate between terms rather than use a single term.⁸⁶ In explanation, Withers and Grout shift between “virtual archive” and “digital archive” in their analysis of a Web-based map archive.⁸⁷ This resulted from, as they state, “We were faced – in truth, more in hindsight than as we proceeded – not with the issues of archives in a ‘post-custodial’ world but, rather, with a ‘multi-custodial’ and, even, a ‘supra-custodial’ world.”⁸⁸

Some of the definitions and uses are broad in scope, such as “the content and services that archival repositories provide to users via the Internet.”⁸⁹ Similarly, according to Galloway, “Digital archiving,” is “the practice of preserving (long term or indefinitely) authentic digital cultural objects for present and future use.”⁹⁰ Some consider digital archives to be collections of born-digital records, while others state, “The ultimate goal of the institution, therefore, is to create hybrid collections – paper, born-digital, and digitized records from the same creating source that are all described in an integrated finding aid.”⁹¹

Another interesting divide is between whether digital archives must follow archival principles. Coleman suggests that, “While the items that comprise an archival collection can vary...these items are required to be stored and displayed with access provided to them based on the original order in which they were created and acquired.”⁹²

⁸⁶ Chandler, “Building Digital Collections at the OAC”; Monks-Leeson, “Archives on the Internet”; Withers and Grout, “Authority in Space?”; Jane Zhang, “Archival Representation in the Digital Age,” *Journal of Archival Organization* 10, no. 1 (2012): 45–68.

⁸⁷ Withers and Grout, “Authority in Space?”

⁸⁸ *Ibid.*, 33.

⁸⁹ Duff et al., “Archivists’ Views of User-Based Evaluation,” 144.

⁹⁰ Galloway, “Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice,” 171.

⁹¹ Tyacke, “Archives in a Wider World,” and Zhang, “Original Order in Digital Archives,” 175–177.

⁹² Coleman, “The Archival and Library Viewpoints of a Collection in a Digital Environment,” 105.

Zhang argues the digital archive must follow two principles: “The first is...respecting provenance and original order; the second is to ensure long-term accessibility of the material.”⁹³ In contrast, Monks-Leeson views “*digital, online, and website archives*,” as created by those “who presumably have little or no grounding in archival theory yet desire to make historical material accessible in digital form.”⁹⁴ This often means creating thematic digital collections of materials, similar to digital libraries. Her discussion of the changing role of archives states:

While in the past *an archive* has referred to a collection of unedited, unannotated material objects, in a digital environment *archive* ‘has gradually come to mean a purposeful collection of surrogates...something that blends features of editing and archiving’... What defines an archive online thus seems to depend on its ability *to archive*, rather than any specificity to its meaning as *an archives*.⁹⁵

In juxtaposition, Samouelian views archival websites as websites of archival institutions “responsible for the long-term preservation of materials.”⁹⁶ Furthermore, Samouelian suggests the difference between a digital collection and digital exhibition is that the former refers to a complete collection while the latter is selective display.

The final point of contention within the archival research is whether digital archives or virtual archives can include material from an external repository. Bearman

⁹³ Zhang, “Original Order in Digital Archives,” 177.

⁹⁴ Monks-Leeson, “Archives on the Internet,” 38. Emphasis original.

⁹⁵ Ibid., 52–53. Emphasis original.

⁹⁶ Samouelian, “Embracing Web 2.0,” 50.

defines virtual archives as “records outside archival custody but under archival control.”⁹⁷

According to Chandler and the Online Archives of California (OAC):

A virtual archive is an electronic grouping of OAC finding aids that collocates and highlights collections sharing a common theme but that are physically dispersed among multiple OAC repositories...While a virtual archive may contain attached images, such images are not required. Proposals for OAC repositories for the creation of virtual archives based on existing OAC finding aids are strongly encouraged.⁹⁸

Westbrook provides a rigid set of definitions starting with highlighting the fact that, “Not all digital collections are virtual collections.”⁹⁹ This is because, in his view, a virtual collection is one created by the user during the use of a digital collection. What others see as a virtual collection, Westbrook calls a *composite collection*, that being “a collection drawn from two or more collections located in the same or different repositories.”¹⁰⁰ He concludes:

There is no real space equivalent to the virtual collection; the collection will be composed of discrete digital objects and digital objects borrowed from their established collection contexts... the virtual collection can be made up of digital items that have never existed together in the same collection.¹⁰¹

⁹⁷ David Bearman, “Virtual Archives.”

⁹⁸ Chandler, “Building Digital Collections at the OAC,” 96.

⁹⁹ Westbrook, “Prospecting Virtual Collections,” 75.

¹⁰⁰ Chandler, “Building Digital Collections at the OAC,” 76.

¹⁰¹ Westbrook, “Prospecting Virtual Collections,” 76–77.

As the above discussion notes, there is no single codified definition of digital archives, nor is there a comprehensive list of generally agreed upon principles or qualities of digital archives (or by any other name). This is most likely due to the continued development of the field, and changing understandings of the role of both born-digital and digitized records within the archival community. The previous discussion provides an overview of the various methods and frameworks of digital archives that must be considered during the discussion of the dissertation findings.

2.2 Digital Archives

Over the past thirty years, digital records began entering archives in ever-increasing numbers, the nature and changing medium of which have caused both great concern and the need to reevaluate archival theory and practice. Some early patrons viewed the digital world as a promising watershed of information. French, for example, noted the research potential of databases of information based on archival information of 18th-century trading and shipping records.¹⁰² Archival practitioners and theorists were more cautious as they saw the onslaught of new technologies streaming past as a threat to the traditional approach for archives. Dryden cautioned, “Digital information is ephemeral. Rapidly changing technology, hardware and software obsolescence, media degradation, and bad records management all threaten the survival of digital information.”¹⁰³ Hedstrom warned, “Digital records will not last long enough to be appraised using conventional practice, as numerous failed attempts to appraise and salvage electronic records, sound recordings, and video tapes from long-inactive systems

¹⁰² French, “Computerizing London’s Eighteenth-Century Maritime Activity.”

¹⁰³ Dryden, “The Open Archival Information System Reference Model,” 217.

have clearly demonstrated.”¹⁰⁴ This ephemeral nature would potentially require archivists to save records “at the moment of creation, or be lost.”¹⁰⁵

The increase in digital materials expands the diversity of archival materials. As Hedstrom states, “The evolving nature of digital documents, broader formulations of memory, and postmodern influences have encouraged me to adopt an open and expansive view of what constitutes records and archives.”¹⁰⁶ The expanse of potential archival materials now includes personal digital photographs housed on Flickr, personal and professional blogs, and email, forcing archivists to find new processing and preservation strategies.¹⁰⁷ Simultaneously, the new digital material and the integration of technology in everyday life lead toward new complications. Cox examines email as the modern letter; additionally, Fredriksson suggests, “the total mixture of official and strictly private information in e-mails,” makes them incredibly difficult from an archival perspective.¹⁰⁸

Recently, in combination with the remains of documentation strategies, archivists explored technology’s potential for broadening collections. Simultaneously, others, such as Nesmith, warn technological advances may, in fact, limit archives’ collecting ability without preemptive measures.¹⁰⁹ Koltun sees the postmodern and technological trends as threatening the foundation of archival practice. She states:

This is the postmodern condition, to chase memory before experience, to focus not on the was, but on the proliferating might be, to rebut teleology, to see life not

¹⁰⁴ Hedstrom, “Archives, Memory, and Interfaces with the Past,” 35.

¹⁰⁵ Koltun, “The Promise and Threat of Digital Options in an Archival Age,” 119.

¹⁰⁶ Hedstrom, “Archives, Memory, and Interfaces with the Past,” 25.

¹⁰⁷ Cal Lee, “Collecting the Externalized Me: Appraisal of Materials in the Social Web,” in *I, Digital*, 201-238.

¹⁰⁸ Cox, “Yours Ever (well, Maybe),” and Fredriksson, “Postmodernistic Archival Science — Rethinking the Methodology of a Science,” 182.

¹⁰⁹ Nesmith, “Seeing Archives.”

as pieced and stitched into an ordered, determinable, and necessary whole, but as unavoidably porous and multiple, subject to particularized, decentred individual perspectives, meshed in continually and rapidly diversifying, never finally coalescing, always contesting discourses.¹¹⁰

Furthermore, she highlights the different nature of digital records within the archives since they are, “the first medium collected by archivists which can be totally dependent on the ‘archiving function’ for its birth, its definition of value, and its continued life.”¹¹¹

The research on digital archives explores three central areas: the nature of digital archives; their current use, and user studies. The exploration of the altering nature of archives in a digital environment leads to interesting questions. Holz, for example, asks, “Are digitization projects just the microfilm of the new millennium? Is the rush to digitize simply a reaction to the funding climate, or is there added value in creating digital instances of existing archival collections?”¹¹² Nimer and Daines, on the other hand, recognize a portion of the digital movement is in response to the “age of instant gratification,” and they stress the need to “reexamine how we present information about our collections online.”¹¹³

Digital archives are changing the method and space of archivist/patron interaction. This is reflective of technology, and as Withers and Grout note:

It is possible to access information about places without being in that place, and for virtual representations to displace real-world encounters and, given claims

¹¹⁰ Koltun, “The Promise and Threat of Digital Options in an Archival Age,” 120.

¹¹¹ Ibid., 123.

¹¹² Holz, “Technologically Enhanced Archival Collections,” 30.

¹¹³ Nimer and Daines, “What Do You Mean It Doesn’t Make Sense?,” 217.

about the relativism of knowledge, for competing claims to authority to be made without, to draw upon Osborne's terms, archival, epistemological, or ethical credibility.¹¹⁴

Furthermore, they see a struggle between users' desires and the limitations of digitization.

They comment:

[...] there remains an emotional and aesthetic relationship between the observer and the original object that the digital image-viewer relationship cannot replicate. The experiences are not the same, and never can be. And yet the digital experience may remain sufficient for all reasonable research-based purposes.¹¹⁵

Rather than the nature of digital archives themselves, other researchers focus on the impact of these collections on archival practice and theory. This includes the added importance of provenance and its relationship with the description of archival materials, as well as the contextual information of record creation and use. Accordingly, Hedstrom notes:

Provenance and the relationship between context and the content of records were considered to be long-standing pillars of archival theory and practice. In the electronic era, they are vital to description, because they provide the key to distinguishing records from non-record material; to understanding why, when,

¹¹⁴ Withers and Grout, "Authority in Space?," 37.

¹¹⁵ *Ibid.*, 45.

and by whom a document was created; and to determining the context in which the record was created, and hence its value and meaning.¹¹⁶

Interestingly, in later work Hedstrom comments on the rise of new decisions, specifically prioritizing item-level description for one digital archival collection over another, and its potential impact on use. She notes, “Materials that are discoverable and accessible remotely will enjoy more use than their physical counterparts, because remote access removes barriers of distance and time.”¹¹⁷

An early advocate on preservation concerns with electronic records, Conway raises significant concerns over validating the quality of digital surrogates from third-party large-scale digitization projects, such as Google Books and HathiTrust.¹¹⁸ Understanding the complexities and labor-intensiveness required for validation by archivists, Conway suggests archivists should “establish user-validated quality metrics for digital surrogates in a very large-scale digital preservation repository of digitized content.”¹¹⁹

Other researchers highlight the opportunity digital archives present for digital repatriation and engaging indigenous communities to better represent their records in the archives. McKemmish, Faulkhead, and Russell, for example, discuss reconciling the research, defined as “[...] a collaborative, co-creative journey, in this case between members of the academy, Indigenous communities and the archival community. It validates multiple sources of knowledge and promotes the use of multiple methods of

¹¹⁶ Hedstrom, “Descriptive Practices for Electronic Records,” 56.

¹¹⁷ Hedstrom, “Archives, Memory, and Interfaces with the Past,” 40.

¹¹⁸ Paul Conway, “Archival Quality and Long-Term Preservation: A Research Framework for Validating the Usefulness of Digital Surrogates,” *Archival Science* 11, no. 3 (2011): 293–309.

¹¹⁹ *Ibid.*, 294.

discovery, implementation and dissemination of knowledge.”¹²⁰ They highlight the need to reincorporate indigenous voices into the digital archives through adapting multiple arrangements and descriptions of the record. Additionally, Ormond-Parker and Sloggett explore indigenous-community archives, particularly digital, and their inclusion within the official record.¹²¹

Christen highlights the importance of using digital archives for repatriation. As she states, “Digital technologies alter repatriation practices by allowing low-cost surrogates of cultural heritage materials to be returned to source communities.”¹²² Specifically, Christen discusses her involvement assisting the development of the Plateau Peoples’ Web Portal Project, which included digital surrogates along with providing “a voice in the curation, narration, and annotation of their materials.”¹²³ The project developed a portal including both scholarly and tribal voices in full detail. Christen notes, “We were not content to simply have a Native ‘comments’ section...Instead, we wanted an integrated metadata scheme that allowed for Native knowledge to be viewed side-by-side with the academic voice.”¹²⁴ Through her positive experience working with the Native peoples and implementing digital repatriation of materials, Christen applied technology to ease the tensions between Native peoples and archives. She concludes, “Opening the collective archival imagination to the diverse needs and heterogeneous hopes of indigenous peoples has the potential to result in a more dynamic and expansive

¹²⁰ Sue McKemmish, Shannon Faulkhead, and Lynette Russell, “Distrust in the Archive: Reconciling Records,” *Archival Science* 11, no. 3–4 (2011): 220.

¹²¹ Lyndon Ormond-Parker and Robyn Sloggett, “Local Archives and Community Collecting in the Digital Age,” *Archival Science* 12, no. 2 (2012): 191–212.

¹²² Christen, “Opening Archives: Respectful Repatriation,” 187.

¹²³ *Ibid.*, 194.

¹²⁴ *Ibid.*, 201.

archive; not a diminished one.”¹²⁵ In looking at the role of the continuum model, online communities, and indigenous populations, Upward, McKemmish, and Reed state:

[...] digital technologies and social networking can support frameworks for the implementation of participatory recordkeeping and archival models (globally and locally), the negotiation of appraisal by records co-creators, the development of meta-metadata schemes that can deal with multiple and parallel provenance and related rights management in current and historical recordkeeping settings, the sharing of recordkeeping and archival spaces, and differentiated access in online cultures.¹²⁶

Simultaneously, the digital technologies “pose challenges to indigenous communities who wish to maintain traditional cultural protocols for the viewing, circulation, and reproduction of these newly animated and annotated cultural materials.”¹²⁷

As with most digital content, archivists have actively participated in the development and implementation of new metadata standards. Vardigan and Whiteman, for example, trace the adaptation of the Open Archival Information System (OAIS) model to the Interuniversity Consortium for Political and Social Research (ICPSR).¹²⁸ Donaldson and Yakel analyze the adoption practices for new metadata standards, such as the Preservation Metadata Implementation Strategies (PREMIS).¹²⁹ Evans, McKemmish,

¹²⁵ Ibid., 210.

¹²⁶ Upward, McKemmish, and Reed, “Archivists and Changing Social and Information Spaces,” 201.

¹²⁷ Christen, “Opening Archives: Respectful Repatriation,” 192.

¹²⁸ Mary Vardigan and Cole Whiteman, “ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context,” *Archival Science* 7, no. 1 (2007): 73–87.

¹²⁹ Devan Ray Donaldson and Elizabeth Yakel, “Secondary Adoption of Technology Standards: The Case of PREMIS,” *Archival Science* 13, no. 1 (2013): 55–83.

and Bhoday discuss the use of automated metadata extraction on accessioned electronic records to provide a fuller contextualization of the records' previous use.¹³⁰

Reviewing the current state of moving-image record digitization, Gracy notes that the cost of high-quality digitization of moving images precludes most repositories from doing anything beyond "creat[ing] an access copy for online distribution that is acceptable for most users."¹³¹ Samouelian, on the other hand, found a large number of repositories already have digital collections (85 of 213 surveyed), while others are "in the process of developing or "hoping to" develop digital collections in the future."¹³²

Digital archives case studies also highlight innovative approaches for displaying archival materials. Watson and Graham report the experiences of creating the CSS Alabama Project, and highlight the use of a "virtual journey" map for user access. This map is:

[...] an exciting experimental method of access that the team hoped would prove appealing as well as geographically instructive, especially to younger users... Users can navigate the route, clicking on the dots to reveal linked log entries, newspaper reports, historical accounts, and illustrations that correspond to events that occurred in the area.¹³³

In dealing with born-digital records, Carroll et al. recognize the importance of the donor's inherent knowledge of the materials "such as how directory structures and file

¹³⁰ Joanne Evans, Sue McKemmish, and Karuna Bhoday, "Create Once, Use Many Times: The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes," *Archival Science* 5, no. 1 (2005): 17–42.

¹³¹ Gracy, "Distribution and Consumption Patterns of Archival Moving Images in Online Environments," 423.

¹³² Samouelian, "Embracing Web 2.0," 57.

¹³³ Watson and Graham, "CSS Alabama 'Digital Collection,'" 129.

naming can map original order,” and the relationships between objects.¹³⁴ The digital archives of Salman Rushdie decided to allow researchers to see a surrogate of Rushdie’s computer to keep the digital structures in place. There is also concern over the preservation of personal digital archives, such as those of fiction writers.¹³⁵

Akmon discusses a case study of one collection’s process of acquiring copyright permission for a digital archive and found the majority of copyright holders granted permission, although the process required significant time.¹³⁶ Dryden found that archives typically follow a more conservative approach to copyright when selecting material for online access.¹³⁷ Only a few other studies consider the impact of copyright concerns on selection for digitization activities.¹³⁸

User studies emerged as a digital archives research focus in recent years since, “We understand little about how the use of archival material changes when accessed in a digital environment.”¹³⁹ Shepard notes the lack of studies discovering “user interest and needs when using digital databases” while discussing online access to archival photographs.¹⁴⁰ Duff and Cherry discuss the altering relationship between archivists and

¹³⁴ Carroll et al., “A Comprehensive Approach to Born-Digital Archives,” 73.

¹³⁵ Becker and Nogue, “Saving-Over, Over-Saving, and the Future Mess of Writers’ Digital Archives.”

¹³⁶ Dharma Akmon, “Only with Your Permission: How Rights Holders Respond (or Don’t Respond) to Requests to Display Archival Materials Online,” *Archival Science* 10, no. 1 (2010): 45–64.

¹³⁷ Jean Dryden, “Copyright Issues in the Selection of Archival Material for Internet Access,” *Archival Science* 8, no. 2 (2008): 123–147.

¹³⁸ Peter J. Astle and Adrienne Muir, “Digitization and Preservation in Public Libraries and Archives,” *Journal of Librarianship and Information Science* 34, no. 2 (2002): 67–79; Barbara Bültmann et al., “Digitized Content in the UK Research Library and Archives Sector,” *Journal of Librarianship and Information Science* 38, no. 2 (2006): 105–122; Gilliland-Swetland, “An Exploration of K-12 User Needs for Digital Primary Source Materials.”

¹³⁹ Duff et al., “Archivists’ Views of User-Based Evaluation,” 147.

¹⁴⁰ Shepard, “Digitizing a Photographic Collection in a Midsize Repository,” 71.

patrons and the need for “more formal evaluation studies to ensure their services and systems meet users’ needs.”¹⁴¹

User studies of digital archives highlighted users’ lack of resource and terminology knowledge. One examination highlighted the lack of discoverability for digital archives. As Allison-Bunnell, Yakel, & Hauck note:

Finally, it is a given that researchers want more materials available on-line. Yet, few of the subjects had used any of the sites in this experiment. This raises the issue that researchers are not aware of many of the sites that do exist, and that there is no one place to go to search all of the archival materials online, nor even any union list of sites. Thus, researchers are not taking full advantage of the existing online archival materials.¹⁴²

In introducing several case studies, Yakel notes the similarity of results and the conclusions that, “Researchers have trouble with archival terminology and are unfamiliar with the hierarchical and provenance-based organization of archives and the search processes in archives.”¹⁴³

Gilliland-Swetland highlighted the need for conducting user studies and then considering the results while planning digital archives, specifically for decisions of what to digitize, the metadata needed for access, and interface design consideration. Unfortunately, archivists are not doing this; she notes, “Instead they are developing

¹⁴¹ Duff and Cherry, “Archival Orientation for Undergraduate Students,” 499.

¹⁴² Allison-Bunnell, Yakel, and Hauck, “Researchers at Work,” 97.

¹⁴³ Elizabeth Yakel, “Balancing Archival Authority with Encouraging Authentic Voices to Engage with Records,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 76–77.

individual digital access initiatives that are rarely fully articulated, systematized across repositories, nor designed based on an analysis of users and their needs.”¹⁴⁴ In recommending strategies for increasing use of digital materials by K-12 students (based on the findings of her user study) Gilliland-Swetland suggests including feedback mechanisms regarding material type, and allowing students and teachers to “contribute critical annotations of the sources they used...that might provide useful descriptive feedback to other K-12 users and archivists.”¹⁴⁵

Adams examined the types of users accessing electronic records at NARA and identified two primary groups: analysts and fact-finders, which “parallel the general categories of users of analog records.”¹⁴⁶ The fact-finders, often genealogists, utilize the digital archives interface more than physical archives users. As more users encounter archives online, the archivist/user relationship is changing “from an archivist-user interpersonal exchange to a user self-service mode.”¹⁴⁷ In another study of governmental archives, Oliver, Chawner, and Liu found workers in New Zealand distrusted the effectiveness of digital archives and their ability to retrieve records online in the same manner as physical archives.¹⁴⁸

Duff et al. found archivists see understanding user needs as an important aspect for prioritizing digitization activities, stating, “Developing and maintaining digital resources is expensive, and they want to make sure they digitize the material users

¹⁴⁴ Gilliland-Swetland, “An Exploration of K-12 User Needs for Digital Primary Source Materials,” 142.

¹⁴⁵ Ibid., 156.

¹⁴⁶ Margaret O’Neill Adams, “Analyzing Archives and Finding Facts: Use and Users of Digital Data Records,” *Archival Science* 7, no. 1 (2007): 34.

¹⁴⁷ Ibid., 31.

¹⁴⁸ Oliver, Chawner, and Liu, “Implementing Digital Archives.”

want.”¹⁴⁹ As one participant stated, “This is the first generation of putting material online and ‘this is a good time to step back’ and evaluate how well we have done to date (Focus Group 1).”¹⁵⁰ Furthermore, Duff et al. argue, “Listening is not enough. We also need to build a culture of assessment that invites comments and feedback from different types of users, both novice and expert.”¹⁵¹

The digital age has had a profound effect on archival theory and practice, and both adapt to the changing technologies and records. Additional studies of the archives in the digital world consider: the digitization of architectural records and three-dimensional models;¹⁵² the use of data grid technology for digital preservation;¹⁵³ the development of digitization standards;¹⁵⁴ blogs as the contemporary diaries and their preservation concerns;¹⁵⁵ the difficulties of preserving listservs;¹⁵⁶ the issues of copyright in digitization projects;¹⁵⁷ the integration of continuum thinking, parallel provenance, the archival multiverse and pluralism;¹⁵⁸ and many other topics.

¹⁴⁹ Duff et al., “Archivists’ Views of User-Based Evaluation,” 157.

¹⁵⁰ Ibid., 158.

¹⁵¹ Ibid., 163.

¹⁵² William J. Mitchell, “Architectural Archives in the Digital Era,” *American Archivist* 59, no. 2 (1996): 200–204.

¹⁵³ Reagan W. Moore, “Building Preservation Environments with Data Grid Technology,” *American Archivist* 69, no. 1 (2006): 139–158.

¹⁵⁴ Nancy Kunde, “Getting It Done—Collaboration and Development of the Digital Records Conversion Standard,” *American Archivist* 72, no. 1 (2009): 146–169.

¹⁵⁵ Catherine O’Sullivan, “Diaries, On-Line Diaries, and the Future Loss to Archives; Or, Blogs and the Blogging Bloggers Who Blog Them,” *American Archivist* 68, no. 1 (2005): 53–73.

¹⁵⁶ Lisa Schmidt, “Preserving the H-Net Email Lists: A Case Study in Trusted Digital Repository Assessment,” *American Archivist* 74, no. 1 (2011): 257–296.

¹⁵⁷ Maggie Dickson, “Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers,” *American Archivist* 73, no. 2 (2010): 626–636.

¹⁵⁸ Michelle Caswell, “On Archival Pluralism: What Religious Pluralism (and Its Critics) Can Teach Us about Archives,” *Archival Science* 13, no. 4 (2013): 273–292; E. Shepherd, The Archival Education and Research Institute (AERI), and Pluralizing the Archival Curriculum Group (PACG), “Educating for the Archival Multiverse,” *American Archivist* 74, no. 1 (2011): 69–101; Upward, McKemmish, and Reed, “Archivists and Changing Social and Information Spaces.”

With the digital emergence, key members of the archival community are beginning to raise concerns over digital archiving pedagogy and the use of digital archives in education.¹⁵⁹ The fast pace of innovation and technological development has quickly exceeded the educational opportunities. As Duff et al. notes, “Currently, the demand for individuals skilled in the area of digital preservation greatly exceeds the supply.”¹⁶⁰ Richard Pearce-Moses also stresses the need for creative thinking and development of new innovative solutions to the current challenges.¹⁶¹

2.3 Minimal Processing

Just as significant as contextualizing the dissertation research within the digital archives research landscape is an understanding of the particular practical application the project addresses. Archives significantly increased the number of accessioned collections following the introduction of postmodernism within archival theory in the early 1970s. Many repositories began, or expanded, collecting manuscript collections in addition to their traditional roles. Finally, the second half of the twentieth century saw increases in both the number and type of records created. These factors, combined with stagnating or reduced workforces, led to higher percentages of accessioned collections remaining inaccessible to the public and unprocessed. The backlog collections ranged between

¹⁵⁹ Anderson et al., “Teaching to Trust”; Wendy M. Duff et al., “Digital Preservation Education: Educating or Networking?,” *American Archivist* 69, no. 1 (2006): 188–212; Galloway, “Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice”; Maija-Leena Huotari and Marjo Rita Valtonen, “Emerging Themes in Finnish Archival Science and Records Management Education,” *Archival Science* 3, no. 2 (2003): 117–129; Lee and Tibbo, “Where’s the Archivist in Digital Curation?”; Richard Pearce-Moses, “Janus in Cyberspace: Archives on the Threshold of the Digital Era,” *American Archivist* 70, no. 1 (2007): 13–22; Gilliland-Swetland, “An Exploration of K-12 User Needs for Digital Primary Source Materials”; Matthew Lyons, “K - 12 Instruction and Digital Access to Archival Materials,” *Journal of Archival Organization* 1, no. 1 (2002): 19–34.

¹⁶⁰ Duff et al., “Digital Preservation Education,” 188.

¹⁶¹ Pearce-Moses, “Janus in Cyberspace.”

twenty-seven and sixty percent of archival holdings.¹⁶² Despite acknowledging the backlog problem, few archivists suggested concrete solutions.¹⁶³ The problem remained ignored by most until the introduction of the More Product, Less Process (MPLP) processing method.¹⁶⁴

This was not the first time that archivists raised concerns about their backlogs and proposed means for addressing them. In a discussion on the improvement needs of historical societies and archives, Josephson questioned, “How large is the backlog of unsorted material awaiting attention in these depositories and how could that backlog be best attacked and attended to?”¹⁶⁵ The National Archives engaged in a massive reappraisal process during the 1950s, thereby addressing its massive backlog created by the accession “spree” during the depression and war years.¹⁶⁶ Fisher also complained of a “stagnating” backlog and stressed the importance of addressing the rising issue.¹⁶⁷

The shrinking budgets and limited staff of many archives prevented them from gaining headway on reducing backlog. As Gorzalski highlights, repositories began seriously considering the storage and processing costs associated with archives.¹⁶⁸ Maher, for example, emphasized the importance of compiling data for both time and money

¹⁶² Barbara M. Jones, “Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in America’s Research Libraries,” *RBM: A Journal of Rare Books, Manuscripts and Cultural Heritage* 5, no. 2 (2004): 88–105; Mary Ellen Rogan, “The Wilson Project, an Archival Success Story,” *Performance! The Newsletter of the Society of American Archivists’ Performing Arts Roundtable, Fall 2006* (2006), <http://www.archivists.org/saagroups/performart/newsletter/PArtsNewsltr2006fal.pdf>.

¹⁶³ Megan Desnoyers, “When Is It Processed?,” *Midwestern Archivist* 7, no. 1 (1982): 5–23; Jones, “Hidden Collections, Scholarly Barriers”; Paul H. McCarthy, “The Management of Archives: A Research Agenda,” *American Archivist* 51, no. 1/2 (1988): 52–69; Helen W. Slotkin and Karen T. Lynch, “An Analysis of Processing Procedures: The Adaptable Approach,” *American Archivist* 45, no. 2 (1982): 155–163.

¹⁶⁴ Greene and Meissner, “More Product, Less Process.”

¹⁶⁵ Bertha E. Josephson, “How Can We Improve Our Historical Societies?,” *American Archivist* 8, no. 3 (1945): 195.

¹⁶⁶ Wayne C. Grover, “Recent Developments in Federal Archival Activities,” *American Archivist* 14, no. 1 (1951): 3–12.

¹⁶⁷ Barbara Fisher, “Byproducts of Computer Processing,” *American Archivist* 32, no. 3 (1969): 215–223.

¹⁶⁸ Gorzalski, “Minimal Processing.”

spent for the sustainable operation of an archive and providing metrics for those outside of the profession (specifically grant-funding organizations) to use in cost-benefit analyses.¹⁶⁹ Subsequently, the archival literature became littered with metric-based studies, with wide-ranging results. Unfortunately, no metric consensus arose from the studies, as each demonstrated the variable in processing speeds from institution to institution ranging from 3.8 hours per cubic foot to 25.2 hours per cubic foot to an incredible 5.5 days per cubic foot.¹⁷⁰

Although the cost-benefit approach and metric analysis indicated some concern over traditional processing costs, it did not directly address the backlog problem. The limited attention given toward providing solutions focused on the same ideas eventually discussed by Greene and Meissner,¹⁷¹ particularly the need for flexibility on processing depth (although other backlog addressing techniques were also introduced, such as reappraisal, speeding processing through team processing, and an early application of computer processing.¹⁷²

Desnoyers remains one of the earliest research suggesting concrete solutions.¹⁷³ She blamed the backlog problem on archivists' lack of defining standard processing levels leading toward archivists who "strive for an ideal that may not always be practical

¹⁶⁹ William J. Maher, "The Importance of Financial Analysis of Archival Programs," *Midwestern Archivist* 3, no. 2 (1978): 3–24.

¹⁷⁰ Uli Haller, "Variations in the Processing Rates on the Magnuson and Jackson Senatorial Papers," *American Archivist* 50, no. 1 (1987): 100–109; Terry Abraham, Stephen E. Balzarini, and Anne Frantilla, "What Is Backlog Is Prologue: A Measurement of Archival Processing," *American Archivist* 48, no. 1 (1985): 31–44; and Gorzalski, "Minimal Processing."

¹⁷¹ Greene and Meissner, "More Product, Less Process."

¹⁷² Caryn Wojcik, "Appraisal, Reappraisal, and Deaccessioning," *Archival Issues* 27, no. 1 (2002): 151–160; and Richard W. Hite and Daniel J. Linke, "Teaming Up with Technology: Team Processing," *The Midwestern Archivist* 15, no. 2 (1990): 91–97; Fisher, "Byproducts of Computer Processing."

¹⁷³ Desnoyers, "When Is It Processed?"

or appropriate.”¹⁷⁴ The increased demand on archivists’ time (particularly on non-processing tasks) and users’ expectations further complicated matters, thereby creating a system where donors are annoyed their donated collections remain unprocessed, researchers’ frustrations grow with inaccessible collections, and archivists remain at a loss. Through reviewing the current situation, Desnoyers recommends archivists begin viewing “processing as a range of choices among a continuum,” rather than always striving for the ideal.¹⁷⁵ In doing so, the archivist “consider[s] the found state of the collection and the requirements and interests of the donor, the users, the applicable legislation, and the material itself.”¹⁷⁶ Desnoyers’ continuum approach explores each step of processing as well as preservation and identifying privacy concerns, with the archivist analyzing the necessary levels prior to undertaking the action.

Slotkin and Lynch based their recommendations on the experiences of an NEH-funded project for the MIT archives which initially proceeded slowly, forcing a rethinking of the processing model.¹⁷⁷ The reexamination resulted in five premises of processing: each collection requires a different level of processing; collections with high research potential should receive more attention; assuming the collection will not be revisited for further processing in the future; every action must occur according to a plan, rather than automatically; and processing works most efficiently in teams rather than individually.¹⁷⁸ Moreover, the preservation activities would also be flexible based on potential research

¹⁷⁴ Ibid., 6.

¹⁷⁵ Ibid., 8.

¹⁷⁶ Ibid.

¹⁷⁷ Slotkin and Lynch, “An Analysis of Processing Procedures.”

¹⁷⁸ Ibid.

use. McCarthy stressed the need to “break from traditional methods,” and recommended a priority-based system, similar to the triage systems found in hospital settings.¹⁷⁹

Although the advocates of an adjustable or flexible processing method existed, their voices did not resonate with the archival establishment until Greene and Meissner took up the charge introducing the MPLP model at the 2005 SAA conference, and expanded it during the 2006 meetings of the Midwest Archives Conference and the Society of California Archivists.¹⁸⁰

Their conference blitz coincided with the formal publication of MPLP in *The American Archivist*, in which they expounded on the ideas of Desnoyers, McCarthy, and Slotkin and Lynch.¹⁸¹ Greene and Meissner laid out the significant backlog issues (including the staggering number of unprocessed collections and the ensuing access limitation for users) through interweaving multiple prior studies’ statistics.¹⁸² The studies used included: an unpublished survey by the SAA Congressional Papers Roundtable (about 33% of repositories had more than 25% and 13% of repositories had more than 50% of collections in backlog); an unpublished survey of the SAA Manuscript Repository from 2003-2004 (60% of repositories had at least 33%, and 34% of repositories had more than 50% of collections in backlog); and a 1992 study by the National Historical Publications and Records Commission (30% of respondents encountered access problems to unprocessed collections). It is important to note that Greene and Meissner’s use of statistics is often regarded as one of the flaws of their paper

¹⁷⁹ McCarthy, “The Management of Archives,” 63.

¹⁸⁰ Gorzalski, “Minimal Processing.”

¹⁸¹ Desnoyers, “When Is It Processed?”; McCarthy, “The Management of Archives”; and Slotkin and Lynch, “An Analysis of Processing Procedures.”

¹⁸² Greene and Meissner, “More Product, Less Process.”

since many of the studies used had either very small sample sizes or did not provide confidence intervals.¹⁸³

Following a lengthy review of the “inconsistent and even schizophrenic” processing literature, Greene and Meissner provide their “golden minimum” solution through one simple question: “What is the least we can do to get the job done in a way that is adequate to user needs, now and in the future?”¹⁸⁴ Focusing on arrangement and description, MPLP echoes its predecessors, arguing for processing variability, with a default point at series-level arrangement and description while leaving the potential for additional levels of processing on a case by case basis. Additionally, Greene and Meissner stress the need for preservation activities to follow the “golden minimum” principles; specifically that “we will rely on our storage area environmental controls to carry the preservation burden” rather than spending time and resources on removing staples, paper clips and refolding.¹⁸⁵ Finally, the MPLP model suggests all “Unprocessed collections should be presume[d] open to researchers. Period,” thereby alleviating some of the access issues noted in earlier surveys¹⁸⁶

While initially offered as an arrangement and description technique, MPLP extends throughout archival processing, practices, and record formats, including appraisal, reference, electronic records (both born-digital and digitized), photographic collections, and privacy issues. MPLP’s impact on reference remains a major concern for some archivists, particularly the potential for shifting cost from processing directly to

¹⁸³ Carl Van Ness, “Much Ado about Paper Clips: ‘More Product, Less Process’ and the Modern Manuscript Repository,” *American Archivist* 73, no. 1 (2010): 129–145.

¹⁸⁴ Greene and Meissner, “More Product, Less Process,” 240.

¹⁸⁵ *Ibid.*, 251.

¹⁸⁶ *Ibid.*, 252.

reference (thereby negating any savings from applying the MPLP model).¹⁸⁷ Maier, for example, discusses the implementation of MPLP at the American Heritage Center (AHC) processing of 537 collections in 2005.¹⁸⁸ During the following fall, the AHC encountered a drastic rise in reference requests related to the minimally processed collection which caused “the reference staff initially [to become] de facto processors in order to provide patrons with description to supplement that found in the catalog record.”¹⁸⁹ Ultimately, the AHC began creating on-demand content lists for collections with reference requests, thereby continuously developing additional description only when requested. Interestingly, the AHC director stated, “Ironically, this conundrum was evidence of the success of the endeavor, as one of the project’s main goals was to alert potential users to the existence of resources for which there had previously been no description, and thus, no access at all.”¹⁹⁰

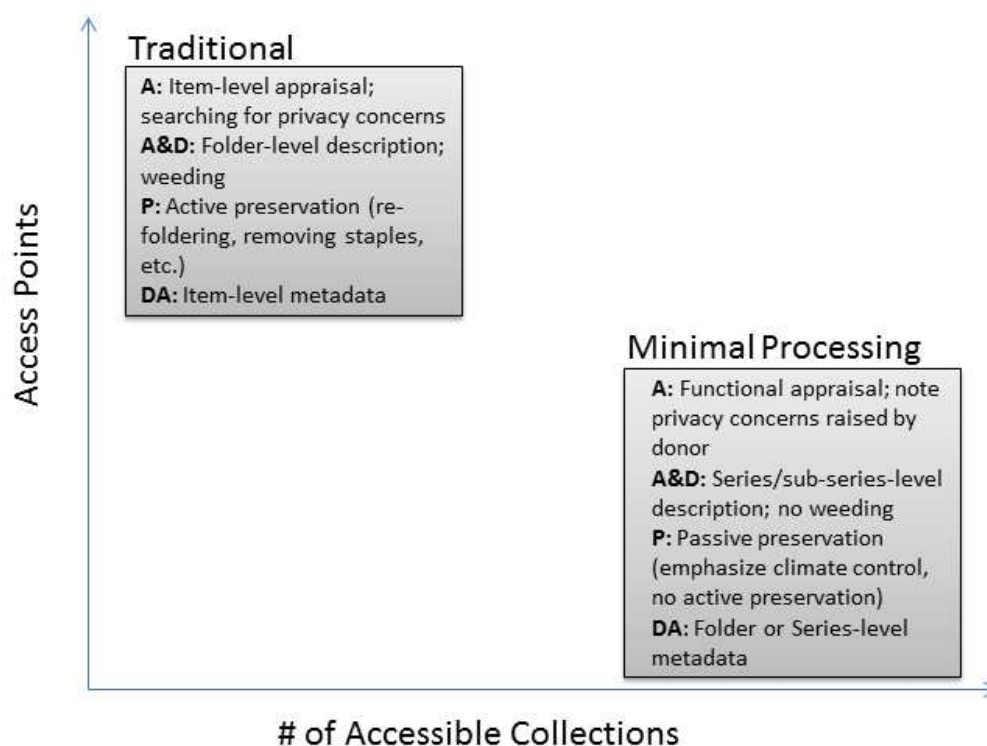
In its original form and application, the minimal processing model shifted processing from micro to more macro practices. Figure 2.1 illustrates the model with specific examples of both traditional and minimal processing from appraisal, arrangement and description, preservation, and digital archives. Adaptation of the minimal processing model also caused a shift in archival access. Traditional processing maintains a high level of access points to the individual collections *already* processed while minimal processing provides increases the number of collections processed.

¹⁸⁷ Stephanie H. Crowe and Karen Spilman, “MPLP @ 5: More Access, Less Backlog?,” *Journal of Archival Organization* 8, no. 2 (2010): 110–133; Greene, “MPLP”; Gorzalski, “Minimal Processing”; Shannon Bowen Maier, “MPLP and the Catalog Record as a Finding Aid,” *Journal of Archival Organization* 9, no. 1 (2011): 32–44.

¹⁸⁸ Maier, “MPLP and the Catalog Record as a Finding Aid.”

¹⁸⁹ *Ibid.*, 41.

¹⁹⁰ *Ibid.*, 42.



Abbreviation key: A: Appraisal, A&D: Arrangement & Description, P: Preservation, DA: Digital Archives

(Adapted from Greene & Meissner, 2005; Greene, 2010; Meissner & Green, 2010)

Figure 2.1 Minimal Processing Model

Through his discussion of further MPLP adaptations, Greene acknowledges its influence on reference services, suggesting that, “At a minimum... [it] requires staff to retrieve more boxes to ensure satisfying the research needs of a patron.”¹⁹¹ Ultimately, though, Greene argues that giving users increased access to previously inaccessible materials far outweighs the increased workload of reference services. Greene dismisses any concern over MPLP’s application to electronic records; specifically referring to Johnson’s discussion.¹⁹² Regarding digitalization efforts, Greene finds no justification for

¹⁹¹ Greene, “MPLP,” 182–183.

¹⁹² Greene, “MPLP.”; Johnson, “Quality or Quantity: Can Archivists Apply Minimal Processing to Electronic Records?”.

an item-level metadata-only approach, citing the work at the University of Wisconsin-Oshkosh and the Smithsonian as examples of digital collections with folder-level or series-level metadata.

Foster discusses the implementation of MPLP on photographic collections through a case study of the University of Alaska Fairbanks (UAF) which applied a minimal processing level on photographs unless noted by user requests or user statistics.¹⁹³ This decision reflects the nature of their users, who are either looking for specific images or all images on a given subject matter. UAF found they could almost never completely satisfy the specific image searchers' expected level of metadata, but could meet the needs of subject searchers through the applied approach. Not only did they experience a rise in user satisfaction; UAF also saw donor relations strengthen.

Several institutions quickly tested the MPLP model following its initial discussion, with mixed results.¹⁹⁴ The adoption of MPLP at Texas Christian University (TCU) involved arrangement- and description-level decisions for each series, each requiring different levels.¹⁹⁵ Strom found the process beneficial, and indicated a continued commitment to the MPLP model. Studies at the University of Montana and Yale University found the MPLP model liberating and drastically increasing the speed of

¹⁹³ Anne L. Foster, "Minimum Standards Processing and Photograph Collections," *Archival Issues* 30, no. 2 (2006): 107–118.

¹⁹⁴ Donna E. McCrea, "Getting More for Less: Testing a New Processing Model at the University of Montana," *American Archivist* 69, no. 2 (2006): 284–290; Jeannette Mercer Sabre and Susan Hamburger, "A Case for Item-Level Indexing: The Kenneth Burke Papers at The Pennsylvania State University," *Journal of Archival Organization* 6, no. 1–2 (2008): 24–46; Michael Strom, "Texas-Sized Progress: Applying Minimum-Standards Processing Guidelines to the Jim Wright Papers," *Archival Issues* 29, no. 2 (2005): 105–112; Christine Weideman, "Accessioning as Processing," *American Archivist* 69, no. 2 (2006): 274–283.

¹⁹⁵ Strom, "Texas-Sized Progress: Applying Minimum-Standards Processing Guidelines to the Jim Wright Papers."

making collections accessible.¹⁹⁶ In discussing the University of Montana's previous state, McCrea notes:

A full-time processor who took eight hours to process each linear foot would just barely keep up with what the archives acquires in a year. Using that same eight hours as an estimate, it would take someone working 40 hours a week, who never got sick, never took vacation, never answered reference questions, and never attended meetings, eleven and a half years to get through our backlog!¹⁹⁷

Following the application of the MPLP approach over two years reduced the average processing time significantly from eight hours per linear foot to two hours.¹⁹⁸

Mercer Sabre and Hamburger object to the series-level application of MPLP at Penn State, stating, "In instances of collections with many disparate items, a series description often can provide little concise information to assist reference staff in service and researchers in discovery."¹⁹⁹ As Crowe and Spilman correctly highlight, the MPLP *does not* limit all processing to series level; rather, it merely suggests series level remain the default processing level.²⁰⁰

More recently, some dispute MPLP's validity and report on its continued divisive nature within archival circles.²⁰¹ Cox argues that archivists do not comprehend the long-

¹⁹⁶ McCrea, "Getting More for Less."; Weideman, "Accessioning as Processing."

¹⁹⁷ McCrea, "Getting More for Less," 285.

¹⁹⁸ McCrea, "Getting More for Less."

¹⁹⁹ Sabre and Hamburger, "A Case for Item-Level Indexing," 44.

²⁰⁰ Crowe and Spilman, "MPLP @ 5."

²⁰¹ Robert S. Cox, "Maximal Processing, Or, Archivist on a Pale Horse," *Journal of Archival Organization* 8, no. 2 (2010): 134–148; Van Ness, "Much Ado about Paper Clips;" and Crowe and Spilman, "MPLP @ 5"; Thomas J. Frusciano, "'The Beat Goes On': MPLP, RDA, Digitization, and Archivist as Historian," *Journal of Archival Organization* 8, no. 3–4 (2010): 169–173.

term costs associated with a purely minimal processing approach.²⁰² He warns, “Small effects operating over a long time can have large consequences.”²⁰³ Furthermore, he states, “If a collection is less well described, less well organized, and less well understood, logic dictates that, all things being equal, it must take longer for archivists to navigate the collection when conducting reference work or when performing any other tasks that make use of the actual materials.”²⁰⁴ Rather than minimal processing being the status quo, Cox argues for a process called maximum processing through which intellectual control begins with similar steps to the minimal processing approach. The major difference, however, is the processing continues after this initial step, as funding allows, through a priority-based system.

Van Ness contends the MPLP is neither a new process nor based on sound statistics.²⁰⁵ He particularly notes the lack of adhering to proper survey methodology, the assumption of a processing metric, the impact of minimal processing on space (weeding and removing duplicates would not occur), and most importantly that backlog is purely a processing problem rather than a combination of appraisal, arrangement, and description. He concludes:

The academic manuscript repository’s preoccupation with minutiae such as paper clips and newspaper clippings is merely symptomatic of a much larger problem.

For the academic library to erase its backlog of historical records, it must do more than streamline its processing procedures. It will have to reverse the current two-to-one ratio of faculty to paraprofessionals and give more attention to the nuts and

²⁰² Cox, “Maximal Processing, Or, Archivist on a Pale Horse.”

²⁰³ Ibid., 139.

²⁰⁴ Ibid.

²⁰⁵ Van Ness, “Much Ado about Paper Clips.”

bolts of processing... Ultimately, the best solution to the backlog problem is not creating one in the first place.²⁰⁶

Based on the criticism and some misinterpretation, this elicited a response from Meissner and Greene reinforcing MPLP's grounding in resource management while providing processing flexibility.²⁰⁷ In doing so, they walk through the various positive reports of MPLP applications including conference workshops, presentations and journal articles prior to entering "the complaints department." Through addressing complaints, Meissner and Greene reiterate the flexible nature of MPLP (it is not a "cookie cutter" approach), dispel the romanticism of item-level description and the "strange mélange of archivists' fears and needs" related to privacy concerns, argue MPLP will not destroy "an important branch of the canon of archival professional literature," find no evidence of an increase in archival theft, and affirm appraisal remains *part* of the backlog problem (but not the sole culprit).²⁰⁸

Overall, application of the minimal processing model (Figure 2.1) increased the amount of publically accessible collections through identifying and using the minimum level of archival involvement and labor throughout processing (appraisal, arrangement & description, and preservation). Additionally, several archives began adapting the model for digital archives through limiting metadata to the folder or series level. Although the model reduces backlogs and increases the number of digital archives available, the minimally processed collections (both analog and digital) offer a reduced number of access points for users.

²⁰⁶ Ibid., 145.

²⁰⁷ Greene, "MPLP."

²⁰⁸ Dennis Meissner and Mark A. Greene, "More Application While Less Appreciation: The Adopters and Antagonists of MPLP," *Journal of Archival Organization* 8, no. 3–4 (2010): 198–216. Emphasis original.

2.4 Postmodernism and Archives

While the dissertation is rooted in the application of minimal processing in a digital archives, the use of social tagging is part of a larger archival postmodern movement. Users bring unique and varying perspectives to each archive, collection and record. Through active engagement with archival materials, and providing tags, the user renews or refreshes the records' context. The participatory archives or Archives 2.0 attempts to integrate these new perspectives into the archival process. The following sections further explore the development and role of postmodernism, the participatory archive, and Archives 2.0 within digital archives.

Howard Zinn infamously caused quite a stir in the 1970s through his lambasting of archivists' reinforcement of the status quo and social control of the political elite. Zinn called on archivists to, "take the trouble to compile a whole new world of documentary material, about the lives, desires, needs, of ordinary people," and, "to begin to play some small part in the creation of a real democracy."²⁰⁹ Zinn's comments, along with others, notably Jacques Derrida, initiated the postmodern movement in archives, and a concerted effort to increase the breadth of voices included in all aspects of archival collecting and practices.²¹⁰ Many archives throughout the past thirty years focused on filling the gaps created by decades of adherence of outdated definitions of records and value through translating postmodernism into new archival practices such as documentation strategy and functional appraisal.²¹¹ Cook describes this as the shift "from the 'nature' residue or

²⁰⁹ Howard Zinn, "Secrecy, Archives, and the Public Interest," *Midwestern Archivist* 2, no. 2 (1977): 14-27.

²¹⁰ Jacques Derrida, *Archive Fever: A Freudian Impression* (Chicago and London: University of Chicago Press).

²¹¹ Helen W. Samuels, "Improving Our Disposition: Documentation Strategy," *Archivaria* 33 (1991-1992): 125-140; Helen W. Samuels, *Varsity Letters: Documenting Modern Colleges and Universities* (Chicago: SAA, 1992).

passive by-product of administrative activity to the consciously constructed and actively mediated 'archivalisation' of social memory."²¹²

Although postmodernism in archival theory remains a debated topic, it is one that is difficult to define, as noted by Cook.²¹³ Unlike early Jenkinsonian archival theory, the postmodern archivist rejects the idealized objectivity of passive record selection and static archival processes in favor of a more dynamic, ever evolving, social memory-focused role. Highlighting the dynamic nature of postmodernism, Nesmith states:

One of the key insights from postmodernism bearing on the reconceptualization of archiving is that it should be seen as an ongoing *process* or action.

Postmodernism suggests that records and archiving, as means of communication, are limited by the various influences and factors which shape them, and their limitations then shape what we can know through them.²¹⁴

Furthermore, the postmodern archive must not try to remove itself from society and its influences by claiming objectivity; rather, as active players or mediators of society. As Heald suggests, stating, "Therefore, we must see ourselves and our institutions as full-fledged members of contemporary society, not as entities that stand outside of it with the aim of documenting it objectively... We must ensure that our focus remains on the records themselves, but we must do so as a willful act of postmodern self-consciousness."²¹⁵

²¹² Cook, "Archival Science and Postmodernism," 4.

²¹³ Cook, "Archival Science and Postmodernism," 5-10. Cook also provides an excellent bibliography and overview of the development postmodern archival research in footnote 13 of his article.

²¹⁴ Tom Nesmith, "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (2005): 261. Emphasis original.

²¹⁵ Carolyn Heald, "Is There Room for Archives in the Postmodern World?" *American Archivist* 59, no. 1 (1996): 101.

The postmodern archives also questions the inherent power dynamic between archivist and users; specifically through the identification of value and ownership of the record.²¹⁶ Cook highlighted this dynamic through discussing necessary changes toward approaching national archives and recommended archivists not limit their collections to the governments themselves but also to the governing process itself. He stated, “‘Governance’ includes being cognizant of the interaction of citizens with the state, the impact of the state on society, and the functions or activities of society itself as much as it does the inward-facing structures of government and its bureaucrats.”²¹⁷ In further discussing the power relationships within archives, Schwartz and Cook state:

Archives have always been about power, whether it is the power of the state, the church, the corporation, the family, the public, or the individual. Archives have the power to privilege and to marginalize. They can be a tool of hegemony; they can be a tool of resistance. They both reflect and constitute power relations.²¹⁸

As part of a dynamic understanding of records, postmodernism captures the struggle to provide and preserve contextual information, since every record can be interpreted in a multitude of ways, and this interpretation may alter over time. As Ketelaar notes, “Once we no longer assume that there is only one reality or meaning or truth, but many, no one better than the other, we can try to find these multiple meanings by interrogating not only the administrative context, but also the social, cultural, political, religious contexts of record creation, maintenance, and use.”²¹⁹ Additionally, Nesmith

²¹⁶ Cook, “Archival Science and Postmodernism,” 9.

²¹⁷ Ibid, 19.

²¹⁸ Joan M. Schwartz and Terry Cook, “Archives, Records, and Power: The Making of Modern Memory,” *Archival Science* 2 (2002):13.

²¹⁹ Eric Ketelaar, “Tacit Narratives: The Meaning of Archives,” *Archival Science* 1, no. 2 (2001): 141.

argues the multiple meanings of records and contextualization can only be known over the course of time, and therefore must be readdressed when necessary.²²⁰

Not only can the archival understanding of a record change over time from the archivist's perspective, but each user brings with himself or herself differing perspectives. The addition (and possible subtraction) of records within an open collection or within a repository may add or remove context and contextualizations.²²¹ Therefore, the user and potential user of archives hold an important role within the postmodern archive. The participatory archive or Archives 2.0 movement can be seen as an extension of postmodernism through an attempt to better integrate the user perspective within archival processes.

2.5 Participatory Archives and Archives 2.0

Shilton and Srinivasan suggest the use of so-called participatory archival applications similar to those suggested by Evans.²²² The participatory archives model engages community members during appraisal, arrangement, and description processes to provide a voice to marginalized communities and increase a sense of empowerment. This concept recently led to new theoretical models of interaction between users and archives. Anderson and Allen, for example, developed the framework for an archival commons, defined as “a space where cultural professionals, researchers, and interested members of

²²⁰ Tom Nesmith, “Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives,” *American Archivist* 65, no. 1 (2002): 36.

²²¹ Verne Harris, “Claiming Less, Delivering More: A Critique of Positivist Formulations on Archives in South Africa,” *Archivaria* 44 (1997): 136.

²²² Shilton and Srinivasan, “Participatory Appraisal and Arrangement for Multicultural Archival Collections.”

the general public could contribute narrative and links among objects of interest held by archives, libraries, and/or museums and *systematically reflect those activities within the primary repository itself.*"²²³

Grounded in Giddens' Structuration Theory, the archival commons develops additional contextual information through user-generated links, both intra-repository and inter-repository. The creation of virtual links between collections allows users to meet their research/use needs through virtually rearranging materials, be it chronologically, thematically, or otherwise. The "new" arrangements and links remain publically accessible and could assist other researchers interested in similar topics. Additionally, this method would benefit instructors since, "No longer would generations of students or groups of students passing through institutions be forced to repeat the laborious process of assembling the same materials for similar purposes either virtually or physically from disparate archival collections."²²⁴

Flinn, one of the leading advocates for participatory archives, argues the interaction between user and record "affect[s] our understanding and knowledge of that archive."²²⁵ Additionally he argues, "Individual and collaborative scholarship and knowledge production are not completely separate modes of working or thinking; they can co-exist and even interact, informing and extending each other."²²⁶ Eveleigh suggests the participatory archives, through engaging more users, could extend archival advocates

²²³ Anderson and Allen, "Envisioning the Archival Commons," 383 Emphasis original.

²²⁴ Ibid., 392.

²²⁵ Flinn, "An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge."

²²⁶ Ibid.

essential in the current state of archives.²²⁷ Huvila views the participatory archives as a method of decentralizing the authority of archives since “Inclusion and greater participation are supposed to reveal a diversity of motivations, viewpoints, arguments and counterarguments, which become transparent when a critical mass is attained.”²²⁸

Moreover, he states:

The motivations for adopting a post-controlled approach and emphasising radical user orientation in a participatory archive by allowing the users to edit actual records is to capture richer descriptions and links between records, to accelerate the process of updating the archive, to engage users to collaborate actively within the archive, and to reduce the need for administrative interventions.²²⁹

Theimer, one of the leading advocates of technological integration, refers to the movement as Archives 2.0 (reflecting the ideas of Web 2.0 and Library 2.0).²³⁰ Defining the term, she states:

Archives 2.0 is an approach to archival practice that promotes openness and flexibility. It argues that archivists must be user centered and embrace opportunities to use technology to share collections, interact with users, and improve internal efficiency...It requires that archivists be active in their communities rather than passive, engaged with the interpretation of their collections rather than neutral custodians, and serve as effective advocates for

²²⁷ Alexandra Eveleigh, “Welcoming the World: An Exploration of Participatory Archives” (presented at the International Conference on Archives, Brisbane, Australia, 2012), 2012, <http://www.gosbook.ru/system/files/documents/2012/11/13/ica12Final00128.pdf>.

²²⁸ Isto Huvila, “Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management,” *Archival Science* 8, no. 1 (2008): 25.

²²⁹ *Ibid.*, 26.

²³⁰ Theimer, *A Different Kind of Web*.

their archival program and their profession. Archives 2.0 is not “something in the future,” but a description of what the majority of archivists believe today.²³¹

Through further expanding her discussion, Theimer reviews the many features of the new 2.0 paradigm including the focus on innovation, flexibility, being technologically savvy, and not becoming obsessed with creating “perfect products.” The technology Theimer champions offers archivists increased engagement with both new and returning users through the use of a variety of Web 2.0 tools, including blogs, wikis, social media, social bookmarking, social tagging, etc. Upward, McKemmish, and Reed note, “Archivists worldwide are beginning to explore the capacity of digital information and new social networking technologies to enhance the accessibility of the traditional custodial archive.”²³²

The motivation for technologically driven outreach includes an appreciation for the modern limitations of archivists. Evans highlighted the perilous modern archival situation of significantly increased collection acquisition combined with fiscal and temporal limitations, suggesting the leveraging of user knowledge through technology to ease the burden. In reference to this model, he states:

Similarly, this model portends an archival system that uses the eyeballs and the intellect of thousands of volunteers—including archival customers, historians, genealogists, students, and others—throughout the world. Acting as partners with archivists, users can do what archivists alone cannot do. Archivists do not have the resources to do item-level description and indexing. But archivists can become

²³¹ Theimer, “What Is the Meaning of Archives 2.0?,” 60.

²³² Upward, McKemmish, and Reed, “Archivists and Changing Social and Information Spaces,” 206.

organizing agents for others to do such work, either independently or as part of social tagging projects.²³³

Ketelaar argues for thinking of the archive as “a dynamic open-ended process,” and suggests the archivists must “connect the memories in our archives with the memories in people’s minds” in order to “make archives into people’s archives.”²³⁴ Gerencser views the interactive nature of Web 2.0 as a better method to reconnect and collaborate with users.²³⁵

Just as digital archives began altering the archivist/user relationship, Palmer and Stevenson argue Archives 2.0 further moves the relationship away from the traditional one-way toward a more dynamic user-driven approach since “attention is now more focused on direct engagement and active interaction with users in online spaces.”²³⁶ Furthermore, Palmer and Stevenson view social media as both promotional and research mechanisms. Jimerson sees the potential for social media to “expand social connections directly with minimal mediation by external experts or gatekeepers.”²³⁷

While many support the Archives 2.0 movement, others raise concerns over the losing of archival authority, and introduction of complexity. As Baxter notes, “Allowing people to interact with information instead of just consuming it can enhance the process, bringing new value to individuals and networks, but it can also muddy the network,

²³³ Max J. Evans, “Archives of the People, by the People, for the People,” *American Archivist* 70, no. 2 (2007): 397.

²³⁴ Eric Ketelaar, “Cultivating Archives: Meanings and Identities,” *Archival Science* 12, no. 1 (2012): 29; and Eric Ketelaar, “Being Digital in People’s Archives,” *Archives & Manuscripts* 31, no. 2 (2003): 12–13.

²³⁵ Gerencser, “New Tools Equal New Opportunities: Using Social Media to Achieve Archival Management Goals.”

²³⁶ Palmer and Stevenson, “Something Worth Sitting for? Some Implications of Web 2.0 for Outreach,” 6.

²³⁷ Randall C. Jimerson, “Archives 101 in a 2.0 World: The Continuing Need for Parallel Systems,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 309.

reducing authority and authenticity and, perhaps, value. It certainly introduces complexity.”²³⁸ Yakel questions the balance between user-generated information and the archival authority.²³⁹ Jimerson highlights the need to think of “Web 2.0 technology [as] a tool, not a goal.”²⁴⁰

In spite of these concerns, Palmer argues for more “risk-taking in respect of crowd-sourcing,” and that “new trust metrics and heuristics will emerge.”²⁴¹ Furthermore, she calls for additional research into the content created by users and how it could be integrated or supplement archival description. Finally, Palmer states, “Users should be treated as *peer collaborators*, intrinsic to the process of meaning-making, rather than outside interlopers (however welcome) who must be kept at arm's length from the authoritative record.”²⁴² Flinn also defends the movement, arguing, “This need not be seen as an attack on professionalism or scholarship. Rather, non-professional participation in online archival activity provides an opportunity to re-think how future professionalism and scholarship might be supported in a more collaborative, inclusive and democratic context.”²⁴³ Eveleigh summarizes both the potential and criticisms alike. She states:

On the one hand then, online user participation is heralded as an opportunity to democratise professional archival practice; promising liberation from the straitjacket of traditional cataloguing practice and promoting the active participation of archives users in co-creating historical meaning. On the other

²³⁸ Baxter, “Going to See the Elephant: Archives, Diversity, and the Social Web,” 286.

²³⁹ Yakel, “Balancing Archival Authority with Encouraging Authentic Voices to Engage with Records.”

²⁴⁰ Jimerson, “Archives 101 in a 2.0 World: The Continuing Need for Parallel Systems,” 305.

²⁴¹ Joy Palmer, “Archives 2.0: If We Build It, Will They Come?,” *Ariadne* 60 (2009), <http://www.ariadne.ac.uk/issue60/palmer>.

²⁴² Ibid. Emphasis original.

²⁴³ Flinn, “An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge.”

hand, participatory culture carries the potential, at least, to subvert not only the hierarchy of the catalogue, but also the power relationships between records, researchers and archivists. User participation initiatives in archives are haunted by a fear that a contributor might be wrong, or that descriptive data might be pulled out of archival context, and that researchers using collaboratively authored resources might somehow swallow all of this without question or substantiation.²⁴⁴

Although the theoretical developments of the Archives 2.0 and postmodernism, as well as their critics, will in time dictate the future directions of the applied research, the majority of current literature on technology's use within archival outreach remains within the applied research arena. Taken as both exploratory research and theoretical experimentation, the following case studies and aggregation of data represent the archival vanguard. The sheer breadth of applications indicates the young nature of the field, and leave room for additional research growth.

Two seminal works explore the potential of a wide variety of Web 2.0 tools through a case study and a survey of existing practice within repositories. Krause and Yakel investigated several Web 2.0 tools and their use within the Polar Bear Expedition Collections providing users several tools for interacting with the collection, including a bookmarking system, user-generated comments, link paths, user profiles, and the traditional browsing and searching features of digital collections.²⁴⁵ Krause and Yakel found the intractability of the finding aid, "transforms it from a static to a dynamic

²⁴⁴ Eveleigh, "Welcoming the World: An Exploration of Participatory Archives," 1.

²⁴⁵ Krause and Yakel, "Interaction in Virtual Archives."

document, an ever-changing resource that provides multidirectional knowledge sharing.”²⁴⁶

Boyer, Cheetham, and Johnson examine using GIS software to manage the City Archives of Philadelphia’s photographic collection.²⁴⁷ Users can access and view photographs of the city on maps, compare the historic images with the modern street view (using Google Street View), comment on images, purchase an image, and notify the archives of potential errors.

Allison-Bunnell, Yakel, and Hauck explored which specific metadata elements provided the most helpful information and were most important for researchers.²⁴⁸ Additionally, the study investigated researchers’ opinions of Web 2.0 tools within digital archives. They found users, “almost always wanted more information about collections and items,” and “they wanted as much detail as possible.”²⁴⁹ This result held true for both textual and non-textual objects alike. Since archivists cannot feasibly describe all digital objects at the item level, “The crucial question becomes not what users want, but what they need.”²⁵⁰ Regarding Web 2.0 tools, Allison-Bunnell, Yakel, and Hauck discovered, “Participants were more interested in taking advantage of information left by other users than in contributing their own information to archival Web sites.”²⁵¹ At the same time, the users thought the archival websites “tended to generate considerably more useful comments than general sites like Flickr or WorldCat,” since there was built in, more

²⁴⁶ Ibid., 308.

²⁴⁷ Boyer, Cheetham, and Johnson, “Using GIS to Manage Philadelphia’s Archival Photographs.”

²⁴⁸ Allison-Bunnell, Yakel, and Hauck, “Researchers at Work.”

²⁴⁹ Ibid., 86–.

²⁵⁰ Ibid., 87.

²⁵¹ Ibid., 92.

dedicated community.²⁵² In looking at how archives and archivists use Facebook and Twitter, Crymble found “Archival organizations overwhelmingly use the services to promote content they have created themselves, whereas archivists promote information they find useful.”²⁵³

In another study, Samouelian analyzed archival websites with digital collections and found a number of them relied on Web 2.0 technologies.²⁵⁴ Samouelian found from follow-up interviews that, “Participants were overwhelmingly positive about using a Web 2.0 application on their repository websites.”²⁵⁵ The archivists suggested users were “the driving force behind the application” of Web 2.0 tools. According to one participant:

[...] we did hear a lot of feedback from people that when they work with images they wanted the ability to add comments, share information—and we certainly are very attentive to that—most of our photographic images come to us with little or no descriptive information, and although there are different types of descriptive information, we wanted an open system that gave and encouraged people to add comments to images and share information so that the next user would have more available information. (Respondent 1).²⁵⁶

Based on her findings, Samouelian views Web 2.0 applications with both strengths and weaknesses. On the one hand, the tools are great for institutional promotion and user engagement; however, the information generated may increase the heavy workload of archivists. She states, “As patrons add comments to blogs and digital images

²⁵² Ibid., 93.

²⁵³ Crymble, “An Analysis of Twitter and Facebook Use by the Archival Community,” 125.

²⁵⁴ Samouelian, “Embracing Web 2.0,” 58.

²⁵⁵ Ibid., 62.

²⁵⁶ Ibid., 63.

or as repositories upload digital images to community sites or even to their own homegrown content management systems, archivists struggle to capture and integrate them into their systems.”²⁵⁷

While the Archives 2.0 movement offers significant potential benefits for both users and archivists, only a handful of institutions are currently integrating or experimenting with these systems. Yakel suggests:

Part of the reason for this may be a wariness of moving away from the traditional relationship between the archivist and the researcher. Another may be the fear of overwhelming responses and actually increasing the work for reference archivists or demands that archives make available more digitized or digitally born materials. Still a few archives and other organizations have begun to let researchers in new and innovative ways.²⁵⁸

Research continues testing different approaches for adapting and utilizing Web 2.0 tools within the archives. For example, Christian and Zanish-Belcher discuss the experience of Iowa State University’s use of YouTube,²⁵⁹ while others highlight applications of Flickr,²⁶⁰ Wikis,²⁶¹ Second Life,²⁶² and blogs.²⁶³ Others explore the

²⁵⁷ Ibid., 65.

²⁵⁸ Yakel, “Inviting the User into the Virtual Archives,” 159.

²⁵⁹ Michele Christian and Tanya Zanish-Belcher, “‘Broadcast Yourself’: Putting Iowa State University’s History on YouTube,” in *A Different Kind of Web: New Connections Between Archives and Our Users*, ed. Kate Theimer (Chicago, IL: Society of American Archivists, 2011), 33–41, <http://works.bepress.com/mchristian/10>.

²⁶⁰ Patrick Peccatte, “Liberating Archival Images: The PhotosNormandie Project on Flickr,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer, trans. Lynne M. Thomas (Chicago: Society of American Archivists, 2011), 148–158.

²⁶¹ Michele Combs, “Wikipedia as an Access Point for Manuscript Collections,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 139–147; Guy Grannum, “Harnessing User Knowledge: The National Archives’ Your Archives Wiki,” in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 116–127; Amy Schindler, “A New Look for

potential of social media's use for using primary sources in the classroom,²⁶⁴ for National History Day research,²⁶⁵ and for outreach.²⁶⁶

The dissertation project is grounded in the minimal processing model, and recognizes the contemporary necessity for a minimal approach. Furthermore, the dissertation puts forth a potentially viable solution for the loss of access points within minimally processed digital archives. Specifically, the supplementation of folder- or series-level metadata with domain expert user-generated tags. Through its application, this solution may begin moving minimally processed collections back toward the high level of access points previously available through traditional processing techniques.

Additionally, the inclusion of social tags within a minimally processed digital archive creates a good adaptation of postmodernism into archival practice. Previous researchers suggested the idea for allowing users to annotate finding aids as a method for integrating a wider variety of interpretations and track their evolution.²⁶⁷ The participatory archives and Archives 2.0 movements encourage the active role of users

Old Information: Creating a Wiki to Share Campus History," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 191–202.

²⁶² Taormina, "The Virtual Archives: Using Second Life to Facilitate Browsing and Archival Literacy."

²⁶³ Malinda Triller, "Double-Duty Blogging: A Reference Blog for Management and Outreach," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 203–212.

²⁶⁴ Jeffrey W. McClurken, "Waiting for Web 2.0: Archives and Teaching Undergraduates in a Digital Age," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 243–254.

²⁶⁵ Tobi Voigt, "Is National History Day Ready for Web 2.0?," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 233–242.

²⁶⁶ Jessica Lacher-Feldman, "Making Friends and Fans: Using Facebook for Special Collections Outreach," in *A Different Kind of Web: New Connections between Archives and Our Users*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 54–64; Medina-Smith, "Going Where the Users Are: The Jewish Women's Archive and Its Use of Twitter."

²⁶⁷ Michelle Light and Tom Hyry, "Colophons and Annotations: New Directions for the Finding Aid," *American Archivist* 65 (2002): 24–41.

within archival description (either officially or as supplemental). Allowing users to tag a digital collection enables them to provide their interpretation of archival records and provides additional contextualization for current and future researchers. Additionally, tagging is a dynamic process that develops and alters over time thereby reflecting the ever-changing interpretation of records.

2.6 Social Tagging

Understanding the placement of the dissertation project within the theoretical and practical needs of archival science and the broader information studies requires an appreciation for the contextualization and development of both the social tagging aspect of Web 2.0 and its applications within digital collections. As such, the following sections outlines the literature of social tagging with an eye toward highlighting the trends, features, and limitations thereof. A more detailed discussion of both archives in the digital world and social tagging follows.

Similar to the development of digital archival theory and practice, the exploration of social tagging begins with a broad background with research on Web-based tagging, mainly for personal use.²⁶⁸ The research shifted to include tagging within traditional information retrieval systems such as databases,²⁶⁹ OPACs,²⁷⁰ and digital libraries.²⁷¹

²⁶⁸ Chufeng Chen, Michael Oakes, and John Tait, "A Location Annotation System for Personal Photos," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, 2006, 726, <http://doi.acm.org/10.1145/1148170.1148339>; Marco Fernandes et al., "Web Annotation System Based on Web Services," in *Proceedings of the International Conference on Next Generation Web Services Practices*, 2005, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1592399; Xin Fu et al., "Annotating the Web: An Exploratory Study of Web Users' Needs for Personal Annotation Tools," in *Proceedings of the American Society for Information Science and Technology*, vol. 42, 2005, <http://onlinelibrary.wiley.com/doi/10.1002/meet.14504201151/abstract>; Cameron Marlow et al., "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read," in *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, 2006, <http://dl.acm.org/citation.cfm?id=1149949>; Catherine C.

Rather than focusing on the systems, many studies examine the tags and taggers themselves. This literature discusses an equally wide variety of topics as above, including taggers and their motivations for tagging,²⁷² how the familiarity of tagging affects the

Marshall, "Toward an Ecology of Hypertext Annotation," in *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space—structure in Hypermedia Systems: Links, Objects, Time and Space—structure in Hypermedia Systems*, HYPERTEXT '98 (New York, NY, USA: ACM, 1998), 40–49, <http://doi.acm.org/10.1145/276627.276632>; P. Jason Morrison, "Tagging and Searching: Search Retrieval Effectiveness of Folksonomies on the World Wide Web," *Information Processing & Management* 44, no. 4 (2008): 1562–1579; Peyman Sazedj and H. Sofia Pinto, "Time to Evaluate: Targeting Annotation Tools," in *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation*, 2005, <http://ceur-ws.org/Vol-185/semAnnot05-04.pdf>; Edith Speller, "Collaborative Tagging, Folksonomies, Distributed Classification or Ethnoclassification: A Literature Review," *Library Student Journal* 2 (2007), <http://www.librarystudentjournal.org/index.php/ljsj/article/view/45>; Victoria Uren et al., "Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art," *Web Semantics* 4, no. 1 (January 2006): 14–28.

²⁶⁹ Morgan Ames and Mor Naaman, "Why We Tag: Motivations for Annotation in Mobile and Online Media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07 (New York, NY, USA: ACM, 2007), 971–980, <http://doi.acm.org/10.1145/1240624.1240772>; Jean-Yves Delort, "Automatically Characterizing Salience Using Readers' Feedback," *Journal of Digital Information* 10, no. 1 (2009), <http://journals.tdl.org/jodi/index.php/jodi/article/view/268>; Jane Hunter, "Collaborative Semantic Tagging and Annotation Systems," *Annual Review of Information Science and Technology* 43, no. 1 (2009): 1–84.

²⁷⁰ Sebastian Chan, "Tagging and Searching--Serendipity and Museum Collection Databases," in *Museums and the Web 2007: Proceedings*, 2007, <http://www.archimuse.com/mw2007/papers/chan/chan.html>; Dion H. Goh Alton Y. K. Chua, "A Study of Web 2.0 Applications in Library Websites," *Library & Information Science Research* 32, no. 3 (2010): 203–211; Jonathan Furner, "User Tagging of Library Resources: Toward a Framework for System Evaluation." (presented at the World Library and Information Congress: 73rd IFLA General Conference and Council, Durban, South Africa, 2007), <http://ifla.queenslibrary.org/iv/ifla73/papers/157-Furner-en.pdf>; Dimitris Gavriliis, Constantia Kakali, and Christos Papatheodorou, "Enhancing Library Services with Web 2.0 Functionalities," in *Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science 5173 (Springer Berlin Heidelberg, 2008), 148–159, http://link.springer.com/chapter/10.1007/978-3-540-87599-4_16; Luiz H. Mendes, Jennie Quiñonez-Skinner, and Danielle Skaggs, "Subjecting the Catalog to Tagging," *Library Hi Tech* 27, no. 1 (2009): 30–41; Tom Steele, "The New Cooperative Cataloging," *Library Hi Tech* 27, no. 1 (2009): 68–77; Jezmyne Westcott, Alexandra Chappell, and Candace Lebel, "LibraryThing for Libraries at Claremont," *Library Hi Tech* 27, no. 1 (2009): 78–81.

²⁷¹ Agosti et al., "Annotation As a Support to User Interaction for Content Enhancement in Digital Libraries"; Bearman and Trant, "Social Terminology Enhancement through Vernacular Engagement"; Matusiak, "Towards User-Centered Indexing in Digital Image Collections"; Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums"; Jennifer Trant, "Tagging, Folksonomy and Art Museums: Early Experiments and Ongoing Research," *Journal of Digital Information* 10, no. 1 (January 12, 2009), <http://journals.tdl.org/jodi/index.php/jodi/article/view/270>; Jennifer Trant, *Tagging Folksonomy and Art Museums: Results of Steve.Museum's Research*, 2009; J. Trant, "Studying Social Tagging and Folksonomy: A Review and Framework," *Journal of Digital Information* 10, no. 1 (2009), <http://journals.tdl.org/jodi/index.php/jodi/article/view/269>; Jason Vaughan, "Insights into the Commons on Flickr," *Libraries and the Academy* 10, no. 2 (2010): 185–214.

²⁷² Luis von Ahn, Ruoran Liu, and Manuel Blum, "Peekaboom: A Game for Locating Objects in Images," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06 (New York: ACM, 2006), 55–64, <http://doi.acm.org/10.1145/1124772.1124782>; Alla Zollers, "Emerging Motivations

quality of tags,²⁷³ the wide range of categories of tags,²⁷⁴ their internal organization,²⁷⁵ and how tags develop.²⁷⁶ Researchers are also reluctant to completely endorse tagging, with some proposing the need for further study of the best utilization of user-generated information.²⁷⁷ More importantly, several studies highlight problems with tagging consistency and use,²⁷⁸ tagging abuse,²⁷⁹ and practitioners' perception of social tagging.²⁸⁰ The literature offers limited potential solutions to consistency issues.²⁸¹

for Tagging: Expression, Performance, and Activism," in *Proceedings of the 16th International World Wide Web Conference*, 2007.

²⁷³ Chei Sian Lee et al., "Tagging, Sharing and the Influence of Personal Experience," *Journal of Digital Information* 10, no. 1 (2009), <http://journals.tdl.org/jodi/index.php/jodi/article/view/275>.

²⁷⁴ Ames and Naaman, "Why We Tag"; Tony Hammond et al., "Social Bookmarking Tools (I): A General Review," *D-Lib Magazine* 11, no. 4 (2005), <http://www.dlib.org/dlib/april05/hammond/04hammond.html>.

²⁷⁵ Pauline Rafferty and Rob Hilderley, "Flickr and Democratic Indexing: Dialogic Approaches to Indexing," *Aslib Proceedings* 59, no. 4/5 (2007): 397–410; Louise F. Spiteri, "The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog," *Information Technology and Libraries* 26, no. 3 (2013): 13–25.

²⁷⁶ Scott A. Golder and Bernardo A. Huberman, "Usage Patterns of Collaborative Tagging Systems," *Journal of Information Science* 32, no. 2 (2006): 198–208; Margaret E.I. Kipp and D. Grant Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices," in *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*, 2006, <http://eprints.rclis.org/archive/00008315/>; Margaret E.I. Kipp, "@toread and Cool: Subjective, Affective and Associative Factors in Tagging," in *Proceedings of the 36th Conference of the Canadian Association for Information Science*, 2008, http://www.cais-acsi.ca/proceedings/2008/kipp_2008.pdf.

²⁷⁷ Louise F. Spiteri, "The Use of Folksonomies in Public Library Catalogues," *The Serials Librarian* 51, no. 2 (2006): 75–89; Robert Fox, "Cataloging for the Masses," *OCLC Systems & Services: International Digital Library Perspectives* 22, no. 3 (2006): 166–172; Katja Šnuderl, "Tagging: Can User-Generated Content Improve Our Services?," *Statistical Journal of the IAOS* 25 (2008): 125–132.

²⁷⁸ Marieke Guy and Emma Tonkin, "Folksonomies: Tidying up Tags?," *D-Lib Magazine* 12, no. 1 (2006), <http://www.dlib.org/dlib/january06/guy/01guy.html>; Kipp and Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices."

²⁷⁹ Georgia Koutrika et al., "Combating Spam in Tagging Systems," in *Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web*, 2007, <http://dl.acm.org/citation.cfm?id=1244420>.

²⁸⁰ Edward Benoit III, "Digital Librarians' Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations," 2012, Manuscript in Preparation; Edward Benoit III, "Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions," 2012, Manuscript in Preparation.

²⁸¹ Ciro Cattuto et al., "Emergent Community Structure in Social Tagging Systems," 11, no. 4 (December 3, 2008): 597–608; Guy and Tonkin, "Folksonomies"; Riina Vuorikari, *Folksonomies, Social Bookmarking and Tagging: The State-of-the-Art*, Special Insight Reports, 2007; Zhichen Xu et al., "Towards the Semantic Web: Collaborative Tag Suggestions," 2006, <http://semanticmetadata.net/hosted/taggingws-www2006-files/13.pdf>.

While major tagging projects exist within both the library and museum worlds with the Library of Congress Flickr²⁸² and Steve.Museum projects,²⁸³ the archival world has not produced similar studies. Small case studies do not analyze the tags produced beyond a quantitative approach.²⁸⁴ While specific cases studies and large-scale studies remain lacking, the respondents of user studies of Web 2.0 tools in general found reluctance to trust un-moderated tags.²⁸⁵ The following section highlights the relevant tagging research focused on tag generation trends and the impact of taggers' motivation(s).

Social taggers' motivation affects the type and quantity of tags in different ways. Zollers concluded that expression, performance and activism as major motivational influences, although different tagging systems attract them in differing proportions.²⁸⁶ Another study concludes users' familiarity with tagging itself may affect the quality of tags produced.²⁸⁷ Ames and Naaman indicate authors are more motivated to tag their own documents.²⁸⁸ Finally, Hammon et al. note:

There is a range from a 'selfish' tagging discipline, where the users are primarily tagging their own content for their own retrieval purposes, right through to a more

²⁸² Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project*; Zinkham and Springer, "Taking Photographs to the People: The Flickr Commons Project and the Library of Congress."

²⁸³ Bearman and Trant, "Social Terminology Enhancement through Vernacular Engagement"; Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums"; Trant, "Tagging, Folksonomy and Art Museums"; Trant, *Tagging Folksonomy and Art Museums: Results of Steve.Museum's Research*; Trant, "Studying Social Tagging and Folksonomy."

²⁸⁴ Tiah Edmunson-Morton, "Talking and Tagging: Using CONTENTdm and Flickr in the Oregon State University Archives," *The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience*, June 19, 2009, <http://interactivearchivist.archivists.org/case-studies/flickr-at-osu/>.

²⁸⁵ Allison-Bunnell, Yakel, and Hauck, "Researchers at Work"; Joyce Celeste Chapman, "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids," *Journal of Archival Organization* 8, no. 1 (2010): 4–30.

²⁸⁶ Zollers, "Emerging Motivations for Tagging: Expression, Performance, and Activism."

²⁸⁷ Lee et al., "Tagging, Sharing and the Influence of Personal Experience."

²⁸⁸ Ames and Naaman, "Why We Tag."

‘altruistic’ tagging discipline, where the user is tagging others’ content for yet others to retrieve.²⁸⁹

Agosti and Ferro see tags as “very broad spectrum, because they range from explaining and enriching an information resource with personal observations to transmitting and sharing ideas and knowledge on a subject.”²⁹⁰ Despite its breadth, Agosti and Ferro developed a complex model for describing the nature of tags. Peters provides an excellent overview and analysis of the literature to date including several models for tags and tagging behavior.²⁹¹ Gupta et al. consolidated the major themes developed over a decade of research on tags and taggers in their survey of different statistical methods used to analyze tags.²⁹² Their literature survey lays the foundation for the hypotheses of research question 1(b), specifically, the participants’ opinions regarding what they considered while creating tags (H₁₀-H₁₄). Gupta et al. identify ten tagging motivations including future retrieval. They state, “Users can tag objects aiming at ease of future retrieval of the objects by themselves or by others.”²⁹³ Furthermore, they stress the use of tags as content description regardless if the future audience is known.²⁹⁴ Finally, Sen et al. state taggers base their tags on personal tendencies (their previous tags) and community influence (other users’ tags).²⁹⁵

²⁸⁹ Hammond et al., “Social Bookmarking Tools (I).”

²⁹⁰ Maristella Agosti and Nicola Ferro, “A Formal Model of Annotations of Digital Content,” *ACM Transactions on Information Systems* 26, no. 1 (2007): 3:2.

²⁹¹ Isabella Peters, *Folksonomies: Indexing and Retrieval in Web 2.0*, trans. Paul Becker (Berlin: Walter de Gruyter, 2009).

²⁹² Manish Gupta et al., “Survey on Social Tagging Techniques,” *ACM SIGKDD Explorations Newsletter* 12, no. 1 (2010): 58-72.

²⁹³ Ibid., 59.

²⁹⁴ Ibid.

²⁹⁵ Shilad Sen et al., “Tagging, Communities, Vocabulary, Evolution,” in *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (New York: ACM, 2006): 180-190.

The majority of research on Web-based systems examines how tags develop or the types of tags. Kipp and Campbell, for example, found tags often develop the same concepts as traditional indexing.²⁹⁶ The quick emergence of consistent tags (those with high frequencies) and the typical inconsistencies shared with multiple indexers, with the addition of spelling, grammar, and synonym errors, show a relationship with common index terms.²⁹⁷ This study also indicates some anomalies which differentiated tags from index terms. Kipp examined this finding further in an additional study, concluding tags often depict emotion, tasks (such as the tag toread) or time.²⁹⁸

Golder and Huberman explored tagging patterns of Delicious and found, despite the overall variety of tags and taggers, some patterns do emerge.²⁹⁹ Similar to Kipp, Golder and Huberman concluded many of the tags were personal in nature, yet still provided some useful information for other users, such as the tag “funny,” which marked a source as personally funny, but which others might find humorous.³⁰⁰ Other examinations of the nature of tags address their inherent inconsistencies, offering potential solutions. Guy and Tonkin, for example, suggest, “Interface changes can be made to discourage certain practices” as well as system suggested common tags to promote consistency.³⁰¹

Perhaps the most promising tagging applications focus on digital collections, and many of these studies are being conducted by practitioners rather than researchers. For

²⁹⁶ Kipp and Campbell, “Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices.”

²⁹⁷ Ibid.

²⁹⁸ Kipp, “@toread and Cool: Subjective, Affective and Associative Factors in Tagging,” 2008.

²⁹⁹ Golder and Huberman, “Usage Patterns of Collaborative Tagging Systems.”

³⁰⁰ Kipp, “@toread and Cool: Subjective, Affective and Associative Factors in Tagging,” 2008; Golder and Huberman, “Usage Patterns of Collaborative Tagging Systems.”

³⁰¹ Guy and Tonkin, “Folksonomies,” 12.

example, the Library of Congress' pilot study examined Flickr to further develop its digital image metadata and an art museum project.³⁰² Bearman and Trant found that "Museum documentation seldom satisfies the on-line access needs of the broad public, both because it is written using professional terminology and because it may not address what is important to—or remembered by—the museum visitor."³⁰³ Additionally, Bearman and Trant highlighted the "profusion of words" which could be used for description of objects and would be "desirable to provide 'keyword' access."³⁰⁴

The internal organization of tags remains a highly debated topic with research indicating a chaotic environment desperately in need of control.³⁰⁵ Other studies suggest user-generated tags conform to the standards of the National Information Standards Organization guidelines.³⁰⁶ The problems of using uncontrolled vocabulary remain one of the central concerns with either integrating folksonomies into metadata or using them as outright indexes. Matusiak examined this issue from a practitioner's perspective and reiterated the unsolved access need for images in digital collections.³⁰⁷ Through her comparison of images in a digital library and in the commercial site Flickr, Matusiak concluded social tagging is not "a simple or miraculous solution to many complex issues inherent in image description."³⁰⁸ Rather than replacing traditional metadata descriptions of images, she recommends the use of tagging as supplemental descriptions. Agosti et al. explored the integration of user-generated information within a digital library interface as

³⁰² Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project*; Zinkham and Springer, "Taking Photographs to the People: The Flickr Commons Project and the Library of Congress," and Trant, *Tagging Folksonomy and Art Museums: Results of Steve.Museum's Research*.

³⁰³ Bearman and Trant, "Social Terminology Enhancement through Vernacular Engagement."

³⁰⁴ Ibid.

³⁰⁵ Rafferty and Hilderley, "Flickr and Democratic Indexing."

³⁰⁶ Spiteri, "The Structure and Form of Folksonomy Tags."

³⁰⁷ Matusiak, "Towards User-Centered Indexing in Digital Image Collections," 286.

³⁰⁸ Ibid., 294.

an enhancement of existing metadata.³⁰⁹ Another approach masks the tag-generating process within a game environment matching terms with images.³¹⁰

2.7 Social Tagging in Digital Libraries

The minority of in-depth digital collection studies include two major projects: the Steve.Museum project led by the Metropolitan Museum of Art and the Library of Congress Flickr project.³¹¹ A significant corpus of literature regarding the use of Flickr began developing following the Library of Congress Flickr project. These studies continued exploring the nature of tags,³¹² proposed methodological metrics,³¹³ highlighted case studies,³¹⁴ explored the experiences of The Commons' participating

³⁰⁹ Agosti et al., "Annotation As a Support to User Interaction for Content Enhancement in Digital Libraries."

³¹¹ Bearman and Trant, "Social Terminology Enhancement through Vernacular Engagement"; Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums"; Trant, "Tagging, Folksonomy and Art Museums"; Trant, *Tagging Folksonomy and Art Museums: Results of Steve.Museum's Research*; Trant, "Studying Social Tagging and Folksonomy;" and Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project*; Zinkham and Springer, "Taking Photographs to the People: The Flickr Commons Project and the Library of Congress."

³¹² Besiki Stvilia and Corinne Jörgensen, "User-Generated Collection-Level Metadata in an Online Photo-Sharing System," *Library & Information Science Research* 31, no. 1 (2009): 54–65; Besiki Stvilia and Corinne Jörgensen, "Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr," *Journal of the American Society for Information Science and Technology* 61, no. 12 (January 1, 2010): 2477–2489; EunKyung Chung and JungWon Yoon, "Categorical and Specificity Differences between User-Supplied Tags and Search Query Terms for Images: An Analysis of Flickr Tags and Web Image Search Queries," *Information Research* 14, no. 3 (2009), <http://www.informationr.net/ir/14-3/paper408.html>; Abebe Rorissa, "A Comparative Study of Flickr Tags and Index Terms in a General Image Collection," *Journal of the American Society for Information Science and Technology* 61, no. 11 (2010): 2230–2242; Oded Nov, Mor Naaman, and Chen Ye, "Analysis of Participation in an Online Photo-Sharing Community: A Multidimensional Perspective," *Journal of the American Society for Information Science and Technology* 61, no. 3 (2010): 555–566.

³¹³ Andrew Cox, Paul Clough, and Stefan Siersdorfer, "Developing Metrics to Characterize Flickr Groups," *Journal of the American Society for Information Science and Technology* 62, no. 3 (2011): 493–506.

³¹⁴ Paul Gahan, "Social Networking, the Swindon Collection," *Multimedia Information and Technology* 36, no. 4 (2010): 25–27; Peggy Garvin, "Photostreams to the People: The Commons on Flickr," *Searcher* 17, no. 8 (2009): 45–49.

institutions,³¹⁵ and compared the tags of the Library of Congress with other Flickr-based institutions.³¹⁶

Art museums represent one of the largest potential digital images distributors and, therefore, require significant improvements within image retrieval systems. For four years, Trant worked with the Metropolitan Museum of Art (MMA) in New York City (and eventually with a vast coalition of art museums in the United States) investigating the potential of social tagging in the art museum community. Since most art museums follow specific internal description standards containing various jargon, general untrained users cannot readily access specific items without prior knowledge of their identifying characteristics (such as accession number, artist, medium, etc.). Additionally, many artistic works' titles do not clearly describe the images contained within. Both issues limit user discovery of new-to-them pieces of art, therefore limiting the educational potential of the institution.

The growth of Flickr-based research increased tremendously following the 2008 Library of Congress Flickr project.³¹⁷ Stvilia and Jörgensen explored the use and nature of photosets on Flickr (not including the Commons).³¹⁸ Relating to tagging, they state, “users did not usually tag individual photos and that the photoset or group metadata were often the only metadata associated with those photos.”³¹⁹ Alternatively, Chung and Yoon

³¹⁵ Vaughan, “Insights into the Commons on Flickr.”

³¹⁶ Benoit III, “Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions.”

³¹⁷ Springer et al., *For the Common Good: The Library of Congress Flickr Pilot Project*; Zinkham and Springer, “Taking Photographs to the People: The Flickr Commons Project and the Library of Congress.”

³¹⁸ Stvilia and Jörgensen, “User-Generated Collection-Level Metadata in an Online Photo-Sharing System.”

³¹⁹ *Ibid.*, 64.

related user-generated tags with query terms used for image searches, finding differences within the specificity of tags versus the query terms.³²⁰

The Flickr-based research continued the trend toward exploration of the nature and similarities/differences between social tags and index terms. Rorissa, for example, compared tags from Flickr images to the index terms of the University of St. Andrews Library Photographic Archive.³²¹ He concluded the tags and index terms are significantly different, and should be used in collaboration for retrieval purposes. Specifically looking at the Library of Congress photo-stream on Flickr, Stvilia and Jörgensen suggest using tag-based folksonomies may “help in vocabulary translation and increase the robustness of traditional [knowledge organization systems] to changes in user expertise, task, and culture.”³²² Nov, Naaman, and Ye explored the nature of the users rather than the tags, finding the long-term users share less photos than new users, while providing more tags.³²³

Although the applications of social tagging within digital collections remains limited, the existing research indicates significant potential. Within a controlled context (applying some of the filtering mechanism discussed earlier), tags give users additional access points to the collections. These new access points typically offer perspectives on items not typically included within official metadata, such as general descriptors (i.e., color, shape, etc.) or more thematic terms. Systems that allow users to sign in could provide personal tracking of interesting or relevant items within the collections.

³²⁰ Chung and Yoon, “Categorical and Specificity Differences between User-Supplied Tags and Search Query Terms for Images.”

³²¹ Rorissa, “A Comparative Study of Flickr Tags and Index Terms in a General Image Collection.”

³²² Stvilia and Jörgensen, “Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr,” 2487.

³²³ Nov, Naaman, and Ye, “Analysis of Participation in an Online Photo-Sharing Community.”

2.8 Social Tagging in Digital Archives

Social tagging within digital archives remains controversial. No matter the technical term, social tagging, user-generated indexing, or user-generated metadata offers users the ability to engage collections on a very personal level, and may increase access points. The reliability and authority of the metadata decrease, however, since the metadata is no longer strictly controlled. For example, Anderson and Allen view tagging, and other Web 2.0 tools, as promising since they “allow users to contribute their knowledge or expertise actively to a project, thereby shaping the interpretation and ensuring cultural meaning.”³²⁴

The archival world has not produced a similar study to the Library of Congress Flickr or Steve.Museum projects. Even at a small scale, only limited literature currently exists. One such study of the Oregon State University Archives on Flickr merely shows the quantitative information, and does not engage the users’ experience or linguistically analyze the tags produced through coding.³²⁵

Bak argues against archives’ use of third-party Web 2.0 systems such as Flickr, stating those which do are following the crowd “without a thought for the loss of value to their own records.”³²⁶ Additionally, he notes the user-generated metadata are “key to the continuing evolution of archival notions of record creation and provenance.”³²⁷ Bak states:

³²⁴ Ibid., 400.

³²⁵ Edmunson-Morton, “Talking and Tagging: Using CONTENTdm and Flickr in the Oregon State University Archives.”

³²⁶ Greg Bak, “Continuous Classification: Capturing Dynamic Relationships among Information Resources,” *Archival Science* 12, no. 3 (2012): 310.

³²⁷ Ibid.

By enabling—and capturing—the mashing, tagging, listing, linking, embedding, blogging, sharing, “liking” (and so on) of records within a recordkeeping or archival system by any user, archives could continue to accumulate metadata that would underwrite a much more sophisticated understanding of records use and repurposing. This, in turn, would feed back into the recordkeeping or archival system to support ever more sophisticated, accurate and user-friendly resource discovery and use.³²⁸

Andreano highlights the potential of social tagging within film archives that can be difficult to access since many archival collections remain poorly described.³²⁹ Although acknowledging the limitations of non-controlled vocabulary, Andreano views the benefits of natural language and “the possibility of serendipitous discovery” as outweighing the limitations since “it is also a relatively cheap and easy way for archives to provide content description.”³³⁰ Yakel highlights a successful implementation of social tagging at the Hague. In her study, “In several cases, multiple visitors have provided increasingly detailed information or corrected the official descriptions.”³³¹

Allison-Bunnell, Yakel, and Hauck found users open to relying on tags if no other item-level description is available; however, the users also questioned the reliability of the tags. Interestingly, “at least one participant felt that the onus was on the other site visitors and not the archivist to vet crowd-sourced information.”³³²

³²⁸ Ibid., 313.

³²⁹ Kevin Andreano, “The Missing Link: Content Indexing, User-Created Metadata, and Improving Scholarly Access to Moving Image Archives,” *The Moving Image* 7, no. 2 (2007): 82–99.

³³⁰ Ibid., 96.

³³¹ Yakel, “Inviting the User into the Virtual Archives,” 161.

³³² Allison-Bunnell, Yakel, and Hauck, “Researchers at Work,” 95–.

Zarro and Allen prefer the inclusion of user comments over tags since commenting allows for a fuller description than a single word or two.³³³ Although discussing commenting, Zarro and Allen suggest self-moderation may be enough of a control mechanism, “with some threshold of ‘thumbs-up’ points needed for a particular comment to be considered trusted.”³³⁴ Furthermore, Townsend recognizes the importance of tagging and other Web 2.0 applications for building and/or strengthening the archivist/user relationship.³³⁵ Although noting, “[...] many academic users would need to be convinced about the long-term value of giving back to the archives,”³³⁶ Townsend argues that archives should cast a wider net toward non-traditional users and communities since “drawing in users to participate in the development of metadata and the process of tagging can potentially extend your staffing resources while leveraging their interest and specific knowledge.”³³⁷ Finally, Townsend suggests opening collections to tagging, and increasing the number of digital archives available will provide evidence for future budget and funding meetings.

2.9 Tagging Issues and Limitations

Social tagging is not without problems. Several researchers discuss the entropic nature of tags and tagging systems, such as variability within spellings, punctuation, and

³³³ Michael A. Zarro and Robert B. Allen, “User-Contributed Descriptive Metadata for Libraries and Cultural Institutions,” in *Research and Advanced Technology for Digital Libraries*, ed. Mounia Lalmas et al., Lecture Notes in Computer Science 6273 (Berlin: Springer, 2010), 46–54, http://link.springer.com/chapter/10.1007/978-3-642-15464-5_7.

³³⁴ *Ibid.*, 53.

³³⁵ Townsend, “Old Divisions, New Opportunities: Historians and Other Users Working with and in Archives.”

³³⁶ *Ibid.*, 224.

³³⁷ *Ibid.*, 220.

compound tag creation.³³⁸ Although Mathes and Golder, as well as Huberman, observed distribution patterns within tags, Kipp and Campbell discovered the patterns do not necessarily always exist, making temporal judgments of tag generation difficult.³³⁹ Additionally, pairs of tags for a given item do not always reflect a relationship, such as synonyms, narrow terms, or broader terms.³⁴⁰ Social tags can also replicate information already provided. In an initial analysis of YouTube tags, Jeong found a high rate (46%) of tags were already included in the titles.³⁴¹ Analysis of a larger sample increased the rate to 52.93% with 54.97% of words in either the title or description also used as tags.³⁴²

Digital librarians remain reluctant to allow tags and other user-generated content within their collections.³⁴³ While they are concerned with possible tag irregularities (i.e., misspellings, compound tag construction, etc.), profanity or spam issues are most troubling, although occurrences of profanity within tagging, such as Flickr are extremely rare.³⁴⁴ Koutrika et al. highlight two related trends within tagging spam, specifically the creation of malicious tags intended to misdirect either a user or the system and so-called

³³⁸ Guy and Tonkin, "Folksonomies"; Kipp and Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices."

³³⁹ Adam Mathes, "Folksonomies: Cooperative Classification and Communication through Shared Metadata" (2004), <http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf>; Golder and Huberman, "Usage Patterns of Collaborative Tagging Systems."; and Kipp and Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices."

³⁴⁰ Ibid.

³⁴¹ Wooseob Jeong, "Does Tagging Really Work?" in *Proceedings of the American Society for Information Science and Technology*, vol. 45, 2008, 1–3, <http://onlinelibrary.wiley.com/doi/10.1002/meet.2008.14504503124/abstract>.

³⁴² Wooseob Jeong, "Is Tagging Effective? – Overlapping Ratios with Other Metadata Fields," *International Conference on Dublin Core and Metadata Applications* (2009): 31–39.

³⁴³ Benoit III, "Digital Librarians' Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations."

³⁴⁴ Ibid.; Benoit III, "Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions."

promotional tagging where a content creator applies unrelated but popular tags to an item to increase viewing.³⁴⁵

Some authors have suggested ways to limit user tagging contributions, especially tags that contain profanity and spam. Moreover, some methods have been devised and/or employed that reduce tagging irregularities, that is inconsistencies within the tags. Guy and Tonkin recommend posting best practices or a tutorial for users to view along with a combination of manual and automatic cleaning of existing tags.³⁴⁶ Others suggest displaying popular tags for new items within a collection or database so users can view existing tags, but ultimately allowing users to add any tags they desire.³⁴⁷ Finally, Xu, Fu, Mao, & Sure commend a combination of approaches including real-time algorithms which highlight statistical outlier tags for possible deletion, tag weighting, and manually moderating tags.³⁴⁸ Cattuto et al. used the tagging information from two semantically opposite terms and compared the similarities between the resulting frequencies of terms.³⁴⁹ In doing so, they applied a TF-IDF weight scheme, thus eliminating “the social aspects of tagging encoded in tag frequencies.”³⁵⁰ Through their analysis, the authors indicate the potential for using vector space modeling as a determining method locating “well-defined communities of resources.”³⁵¹

³⁴⁵ Koutrika et al., “Combating Spam in Tagging Systems.”

³⁴⁶ Guy and Tonkin, “Folksonomies.”

³⁴⁷ Vuorikari, *Folksonomies, Social Bookmarking and Tagging: The State-of-the-Art*.

³⁴⁸ Xu et al., “Towards the Semantic Web: Collaborative Tag Suggestions.”

³⁴⁹ Cattuto et al., “Emergent Community Structure in Social Tagging Systems.”

³⁵⁰ Ibid., 601.

³⁵¹ Ibid., 607.

2.10 Expert and Novice Users

Users' expertise levels, both from a domain and system perspective, remain highly associated with their success and experience during information searching and retrieval. Social tagging provides various degrees of search support depending on the user's previous experience with the system and his/her prior subject knowledge. The IR efficiency of users varies based upon the four combinations of expertise (system expert/domain expert [SEDE]; system expert/domain novice [SEDN]; system novice/domain expert [SNDE]; and system novice/domain novice [SNDN]), thereby requiring a review of their associated characteristics.

Some studies on Internet use suggested a high correlation between a user's system knowledge and comfort with information searching and retrieval, while other studies could not confirm such a relationship.³⁵² Marchionini suggests the sharp learning curve of

³⁵² C L Borgman, "The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog," *International Journal of Man-Machine Studies* 24, no. 1 (1986): 47–64; Janette R. Hill and Michael J. Hannafin, "Cognitive Strategies and Learning from the World Wide Web," *Educational Technology Research and Development* 45, no. 4 (1997): 37–64; Christine Jenkins, Cynthia L. Corritore, and Susan Wiedenbeck, "Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise.," *IT & Society* 1, no. 3 (2003): 64–89; Julita Vassileva, "A Task-Centered Approach for User Modeling in a Hypermedia Office Documentation System," *User Modeling and User-Adapted Interaction* 6, no. 2–3 (1996): 185–223; Nigel Ford, David Miller, and Nicola Moss, "The Role of Individual Differences in Internet Searching: An Empirical Study," *Journal of the American Society for Information Science and Technology* 52, no. 12 (2001): 1049–1066; Tricia Jones, "Incidental Learning during Information Retrieval: A Hypertext Experiment," in *Computer Assisted Learning: Proceedings of the Second International Conference*, ed. Hermann Maurer (Berlin: Springer-Verlag, 1989), 235–253; Kushal Khan and Craig Locatis, "Searching through Cyberspace: The Effects of Link Display and Link Density on Information Retrieval from Hypertext on the World Wide Web," *Journal of the American Society for Information Science* 49, no. 2 (1998): 176–182; Gary Marchionini, *Information Seeking in Electronic Environments* (Cambridge: Cambridge University Press, 1995); Ragnan Nordlie, "User revealment—A Comparison of Initial Queries and Ensuing Question Development in Online Searching and Human Reference Interaction," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, <http://dl.acm.org/citation.cfm?id=312624.312618>; Xianhua Wang, Peter Liebscher, and Gary Marchionini, *Improving Information-Seeking Performance in Hypertext: Roles of Display Format and Search Strategy* (College Park: University of Maryland, 1988).

novice users, regarding system use, might preclude definitive system-based efficiency judgment differentials between novice and expert users.³⁵³

System experts tend to apply advanced search features and complex query structures more often than novice users.³⁵⁴ System novices often require more multiple query re-formulations per task than expert users, and even a higher number if the system novice is also a domain novice.³⁵⁵ In the case of Boolean searching, Ford, Miller, and Moss argue the link between successful complex Boolean queries and expert system knowledge “is hardly surprising.”³⁵⁶ Since formulating search queries for Boolean searching requires, relative to Best-match, a greater level of particular technical skill and knowledge, one would expect individuals lacking relevant experience to demonstrate less use of this strategy.”³⁵⁷

Low levels of information literacy and system knowledge lead to bouncing behavior.³⁵⁸ System novices typically employ a “breadth-first pattern of information seeking,” produce a low level of performance, and encounter a high cognitive load.³⁵⁹ Furthermore, Martzoukou states, “Inadequate system knowledge can transform the search for information into a time-consuming process that increases the cognitive load on the

³⁵³ Gary Marchionini, “Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia,” *Journal of the American Society for Information Science* 40, no. 1 (1989): 54–66.

³⁵⁴ Christoph Hölscher and Gerhard Strube, “Web Search Behavior of Internet Experts and Newbies,” *Proceedings of the 9th International World Wide Web Conference on Computer Networks: The International Journal of Computer and Telecommunications Networking* 33, no. 1 (2000): 337–346.

³⁵⁵ Ibid.

³⁵⁶ Nigel Ford, David Miller, and Nicola Moss, “Web Search Strategies and Human Individual Differences: A Combined Analysis,” *Journal of the American Society for Information Science and Technology* 56, no. 7 (2005): 757–764.

³⁵⁷ Ibid., 760.

³⁵⁸ David Nicholas et al., “Characterising and Evaluating Information Seeking Behaviour in a Digital Environment: Spotlight on the ‘bouncer,’” *Information Processing & Management* 43, no. 4 (2007): 1085–1102.

³⁵⁹ Jenkins, Corritore, and Wiedenbeck, “Patterns of Information Seeking on the Web”; A. G. Sutcliffe, M. Ennis, and S. J. Watkinson, “Empirical Studies of End-User Information Searching,” *Journal of the American Society for Information Science* 51, no. 13 (2000): 1211–1231.

user, while increased experience can positively affect the quality of the user's searching tactics.”³⁶⁰

The focus of system and domain experts provides a significant difference as system experts center on precise queries and searching skills; whereas domain experts evaluate the content of retrieved documents in more depth.³⁶¹ Furthermore, during search result evaluation, system experts assess more general elements (i.e., publication date, title, etc.) than the domain specialist.³⁶²

Similar to system expertise, early studies on Web searching indicated high performance of domain experts when compared to their novice counterparts.³⁶³ A high degree of prior domain knowledge allows expert users the ability to create more specific queries (conceptual-wise) than novice users.³⁶⁴ An analysis of domain experts and novices found, “Novices engaged in less effective strategic search behavior... [Experts’] overall searches were rated as more complex, and they incorporated significantly more

³⁶⁰ Konstantina Martzoukou, “A Review of Web Information Seeking Research: Considerations of Method and Foci of Interest,” *Information Research* 10, no. 2 (2005): User experience section, para. 5, <http://www.informationr.net/ir/10-2/paper215.html>.

³⁶¹ Gary Marchionini, Sandra Duggins, and Xia Lin, “Effects of Search and Subject Expertise on Information-Seeking in a Hypertext Environment,” in *Proceedings of the 53rd Annual Meeting of the American Society for Information Science*, 1990, 129–142.

³⁶² Gary Marchionini, “Information-Seeking in Full-Text End-User-Oriented Search Systems: The Roles of Domain and Search Knowledge,” *Library and Information Science Research* 15, no. 1 (1993): 35–69.

³⁶³ Sandrag Hirsh, “How Do Children Find Information on Different Types of Tasks? Children’s Use of the Science Library Catalog,” *Library Trends* 45, no. 4 (1997): 725–745; Ard W. Lazonder, Harm J. A. Biemans, and Iwan G. J. H. Wopereis, “Differences Between Novice and Experienced Users in Searching Information on the World Wide Web,” *Journal of the American Society for Information Science* 51, no. 6 (2000): 576–581; Marchionini, *Information Seeking in Electronic Environments*; Sharon McDonald and Rosemary J. Stevenson, “Navigation in Hyperspace: An Evaluation of the Effects of Navigational Tools and Subject Matter Expertise on Browsing and Information Retrieval in Hypertext,” *Interacting with Computers* 10, no. 2 (1998): 129–142; Swapnesh C. Patel, Colin C. Drury, and Valerie L. Shalin, “Effectiveness of Expert Semantic Knowledge as a Navigational Aid within Hypertext,” *Behaviour & Information Technology* 17, no. 6 (1998): 313–324.

³⁶⁴ Helene A. Hembrooke et al., “The Effects of Expertise and Feedback on Search Term Selection and Subsequent Learning,” *Journal of the American Society for Information Science and Technology* 56, no. 8 (2005): 861–871.

unique terms... than novices.”³⁶⁵ This difference remains based on the user struggling with problem representation.³⁶⁶ The difference between domain experts and novices provides the major comparison group for the dissertation study. As such, it serves as the foundation for the majority of tested hypotheses including: H₁-H₉ and H₁₅-H₁₆.

Overall SEDE users have the most successful information retrieval effectiveness based upon their use of advanced search features, complex queries, highly conceptual search terms, and in-depth content evaluation skills in combination with a well-developed set of searching tactics. Since social tags offer supplementary searching information from the traditional index terms, to which the SEDE user is well accustomed, tags would not provide a significant impact on the SEDE’s IR effectiveness. One possible exception could occur with private tags provided by the SEDE, such as those within a social bookmarking website, as the personal nature associated with these tags extends beyond the traditional IR system.

On the other end of the expertise spectrum, SNDN struggle the most with information retrieval due to their use of simple searching techniques, bouncing behavior, cognitive overloading, basic search terms, and limited evaluation skills. Unlike SEDEs, the SNDN would benefit greatly from the inclusion of social tags within an IR system. The tags provide additional matching terms for the system to match the user with the required information. SNDN users can use individual tags (or a combination thereof) as access points to similar documents. Finally, the application of tag clouds provides a visual representation that may assist SNDNs with query building. As SEDEs and SNDNs

³⁶⁵ Ibid., 867.

³⁶⁶ Michelene T. H. Chi, Robert Glaser, and Marshall J. Farr, eds., *The Nature of Expertise* (Hillsdale, NJ: Erlbaum, 1988).

provide the extremes, the remaining users fall at different points within the spectrum, with all, except for SEDEs, benefitting to some degree from the inclusion of social tags.

2.11 Limitations of Literature

As archives entered the digital era, archival theory and practice struggled to embrace the changing nature of records, expanded collections, and developing best practices for providing online access to archival materials. Researchers focused initial efforts on generating online finding aids and found the new online users were confused with archival terminology, arrangement, and the simple fact that only limited numbers of records were digitally available. Although some digital archives offer item-level description, this remains cost prohibitive, and the trend toward minimal processing theory will most likely limit description to the folder level. This will be acceptable for traditional archival users, but does not meet user demands for increased access points. While some studies explore the potential for increasing user control and flexibility through utilizing Web 2.0 tools, these remain limited and have not been applied to a large number of existing collections. Additionally, users remain skeptical of un-moderated user-generated content.

The social tagging research, as a whole, appears well developed through its exploration of tagging with IR- and Web-based systems, and the nature of tags and taggers. Additionally, the concerns over applications of tagging within traditional controlled vocabulary settings, such as digital collections, are well expressed. What remains unexamined, however, is empirical testing of control mechanisms which address these concerns. Additionally, tagging within digital archives has not received as much

attention within the research, as tagging in digital libraries due to the lack of major tagging projects related to archives.

The dissertation addresses the gaps of both the archival and tagging literature. From an archival perspective, the dissertation provides a possible solution for increasing the access points within minimally processed digital archives within a postmodern framework. This will further develop the Archives 2.0 research with an easy and practical application while addressing the user's demand for more item-level description. Through examining the use of expert user-generated tags, the dissertation also provides a possible quality-control mechanism for the tags requiring limited oversight by the archivist.

CHAPTER THREE: METHODOLOGY

This study focuses on three main research questions and their sub-questions:

RQ1a—What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?; RQ1b—Are there differences between expert and novice users’ opinions of the tagging experience and tag creation considerations?; RQ2a— In what way do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?; RQ2b—Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?; RQ3a—In what way do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users’ search terms in a digital archive?; and RQ3b—Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive? A mixed-methods, quasi-experimental design best addresses these questions by focusing on tag generation for a sample minimally processed digital archive. Table 3.1 provides an overview of the data-collection methods and analysis for each research question.

Table 3.1 Research Questions and Associated Data and Analysis

Research Question	Data Collected	Data Analysis
RQ1a: What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?	Pre-questionnaire	Descriptive statistics
	Tags generated by expert and novice users (at least one tag for 30 items per participant)	Open-coding, descriptive statistics

H ₁ : The number of tags generated in a minimally processed digital archive is affected by a user's domain knowledge.	Tags generated by expert and novice users (at least one tag for 30 items per participant)	Independent-samples t-tests
H ₂ : The number of photographic tags generated in a minimally processed digital archive is affected by a user's domain knowledge.		Independent-samples t-tests
H ₃ : The number of document tags generated in a minimally processed digital archive is affected by a user's domain knowledge.		Independent-samples t-tests
H ₄ : The proportion of tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.		Chi-square tests for association, Phi, and Cramer's V
H ₅ : The proportion of photographic tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.		Chi-square tests for association, Phi, and Cramer's V
H ₆ : The proportion of document tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.	Post-questionnaire	Chi-square tests for association, Phi, and Cramer's V
RQ1b: Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?		Content analysis, descriptive statistics
H ₇ -H ₉ : Expert and novice users' opinions of the tagging experience are different for ease of tagging in general (H ₇); difficulty in tagging documents compared to photographs (H ₈); and difficulty in tagging photographs compared to documents (H ₉).		Mann-Whitney U tests

<p>H₁₀-H₁₄: Expert and novice users' opinions of the considerations for the creation of tags are different for how others would find the item (H₁₀); how the tagger (user) would find the item (H₁₁); the content of the tagged item (H₁₂); the format of the tagged item (H₁₃); and other users' tags (H₁₄).</p>		Mann-Whitney U tests
<p>RQ2a: In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?</p>	<p>Tags generated by expert and novice users (at least one tag for 30 items per participant), unselected metadata from <i>March on Milwaukee</i></p>	Descriptive statistics
<p>RQ2b: Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?</p> <p>H₁₅: The proportion of tags matching unselected metadata is affected by the user's domain knowledge</p>	<p>Comparison of generated tags (by group) to unselected metadata (by record) tables, and comparison of generated tags (by group) to unselected metadata (all records) tables</p>	Chi-square tests for association, Phi, and Cramer's V
<p>RQ3a: In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?</p>	<p>Tags generated by expert and novice users (at least one tag for 30 items per participant), <i>March on Milwaukee</i> query list extracted from server logs</p>	Descriptive statistics
<p>RQ3b: Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?</p>	<p>Comparison of users' query terms and sample collection metadata/tags table</p>	Chi-square test for association, Phi, and Cramer's V
<p>H₁₆: The proportion of tag terms matching users' query log terms is affected by user's domain knowledge</p>		

3.1 Sample Collection

This study uses selections from an existing digital collection to create a sample digital archive for the experiment. The creation of a sample collection derived from an existing collection creates a comfortable setting and interface for participants during the data collection, thereby strengthening the internal validity of the data. The sample collection is primarily used for research question 1. Additional data from the existing collection, specifically metadata not selected for the sample collection and server query logs, is used for the analysis of research questions 2 and 3.

Rather than a random sampling from a single collection, the sample collection uses a critical case-sampling technic. A random sampling would not necessarily include items previously used within the existing digital collection and would therefore limit the amount of existing metadata needed for comparison with the tags terms generated. The critical case approach allows, “the researcher [to] select a limited number of cases that logic or prior experience indicate will allow generalization to the population.”³⁶⁷ For the dissertation project, the collection population under consideration includes all digital archives (as defined earlier). In this case, the selection procedure prioritized the format over content and included a combination of handwritten documents, typed documents, and photographic images.

The sample collection includes thirty selected records from *The March on Milwaukee Civil Rights History Project* (hereafter called *March on Milwaukee*); a University of Wisconsin-Milwaukee Libraries digital collection. *March on Milwaukee* is a curated digital collection containing about 150 objects from thirteen archival collections

³⁶⁷ Gary T. Henry, *Practical Sampling* (Newbury Park, CA: Sage, 1990), 21.

with a wide range of formats including audio, documents (handwritten and typed), photographs, and moving images. *March on Milwaukee* includes archival materials from multiple collections related to the civil rights movement in Milwaukee for the purpose of “mak[ing] Milwaukee’s place in the national struggle for racial equality more accessible, engaging and interactive.”³⁶⁸ The collection has been active since 2010, and received awards from the Wisconsin Historical Records Advisory Board, the Wisconsin Historical Society, the American Association of State and Local History, and the Society of American Archivists.³⁶⁹ The collection also received coverage in local, regional and national media. The dissertation project uses a sample collection extracted from the *March on Milwaukee* for three primary reasons. First, as recognized by media coverage and awarding bodies, *March on Milwaukee* is a well-constructed and popular collection. It provides excellent existing metadata for comparison with the generated tags. Second, the query logs required for data analysis are available and readily obtainable since UW-Milwaukee servers house the collection. Finally, the researcher’s familiarity of *March on Milwaukee*’s subject matter allows him to better analyze the generated tags and concentrate recruitment on target populations if necessary.

March on Milwaukee contains material from thirteen different collections including both personal and organizational records. The personal papers of one of the main leaders of the Milwaukee movement, James Groppi, is included within *March on Milwaukee*, and was selected as the sole source for the sample collection’s records. This particular collection was selected as the sole source for the sample collection’s records

³⁶⁸ *March on Milwaukee Civil Rights History Project*,
<http://collections.lib.uwm.edu/cdm/landingpage/collection/march>.

³⁶⁹ *March on Milwaukee*, “Awards & Media Coverage,”
<http://collections.lib.uwm.edu/cdm/about/collection/march#awardsrec>.

since the collection contain multiple formats of materials. The selected records were equally divided between images and documents with the latter further divided into three groupings (based on the existing arrangement and description of the Groppi Papers): hate mail, support mail, and criticism mail (see Appendix A). Each of the four series/subseries of records was uploaded into a CONTENTdm hosted digital collection as a compound object thereby maintaining the contextual relationship between records within each grouping. Adhering to the aforementioned minimal processing practice, each compound object will only display a shared minimal metadata set (see Table 3.2).

Table 3.2 Sample Collection Minimal Metadata

Title	Groppi Papers, Correspondence, Hate Mail	Groppi Papers, Correspondence, Support Mail	Groppi Papers, Correspondence, Criticism Mail	Groppi Papers, Photographs
Part of Collection	James Groppi Papers, 1956-1978			
Creator	Groppi, James, 1930-1985			
Type (DCMI)	Text			Image
Original Collection	James Groppi Papers, 1956-1978			
Original Item Location	Milwaukee Mss EX. Box 8, Folders 3-6	Milwaukee Mss EX. Box 1, Folders 1-6	Milwaukee Mss EX. Box 5, Folder 6	PH 4983
Original Item Type	Documents			Photographs
Finding Aid	http://digital.library.wisc.edu/1711.dl/wiarchives.uw-whs-mil000ex			
Repository	Archives / Milwaukee Area Research Center. University of Wisconsin-Milwaukee Libraries			
Digital Publisher	University of Wisconsin-Milwaukee Libraries			
Date Digitized	2010			
Digital Format	image/jp2			
Digital Collection	March on Milwaukee - Civil Rights History Project			
Rights	The Wisconsin Historical Society			

3.2 Sample Population

The dissertation project data was generated from sixty participants divided equally through purposive sampling based on domain knowledge of the civil rights movement in Milwaukee.³⁷⁰ The overall population group focuses on the metropolitan Milwaukee area because *March on Milwaukee* would most likely be accessed in the real world by users from the region. Participants were limited to those over eighteen years old; however, no additional exclusion criteria were enforced, ensuring diverse sample population demography.

Participants were recruited through various methods including online postings, flyers, and directed invitations. Since the dissertation project requires both expert and novice users, recruitment methods targeted potential participants from both groups. Online postings on websites, such as Craigslist, were most successful for gathering participants within the novice grouping, while directed invitations were sent to college instructors of Milwaukee history in the local region to pass on to their students.³⁷¹ The researcher leveraged contacts developed from a previous conference on the civil rights movement in Milwaukee, and known researchers of the subject to meet the required thirty experts. Additionally, invitations were sent, and flyers posted at local historical societies and archives to include archival researchers within the participant pool. Participant recruitment continued on a rolling basis, with focused, directed recruitment toward the end, until the required number of participants for each group was met.

³⁷⁰ Barbara M. Wildemuth, "Sampling for Extensive Studies," in *Applications of Social Research Methods to Questions in Information and Library Science*, ed. Barbara M. Wildemuth (Westport, CT: Libraries Unlimited, 2009), 116–128.

³⁷¹ *Craigslist*, <http://www.craigslist.org>.

In order to increase the response rate, and since participation in the study requires a time commitment of about 1.5-2 hours, each participant was compensated \$15 upon their completion of the study. At first glance, the use of financial incentives for study participation raises serious ethical considerations. Of prime concern is whether the incentives themselves unduly influence or coerce participants. In order to judge the effect, according to Singer and Couper, “that the criterion should be whether or not they induce participants to undertake risks they would not be willing to accept without the incentive.”³⁷² Since the dissertation project does not involve significant risk, the incentives are not coercive. Singer and Couper also note that, “if there are only minimal risks in research—that is, risks no greater than those in ordinary life—the size of the incentive becomes irrelevant on ethical grounds.”³⁷³ The dissertation study also meets this criterion.

Another concern regarding incentives is its effect on the makeup of sample populations and its impact on data collected during the study. Cantor, O’Hare, and O’Connor found incentives had no significant effect on the sample demographics.³⁷⁴ Singer and Kulka also comment on data integrity, noting that the evidence “suggests that the quality of responses given by respondents who receive a prepaid or a refusal conversion incentive does not differ from responses given by those who do not receive an

³⁷² Eleanor Singer and Mick P. Couper, “Do Incentives Exert Undue Influence on Survey Participation? Experimental Evidence,” *Of Empirical Research on Human Research Ethics* 3, no. 3 (2008): 3.

³⁷³ *Ibid.*, 7.

³⁷⁴ David Cantor, Barbara C. O’Hare, and Kathleen S. O’Connor, “The Use of Monetary Incentives to Reduce Nonresponse in Random Digit Dial Telephone Surveys,” in *Advances in Telephone Survey Methodology*, ed. James M. Lepkowski et al. (Hoboken, NJ: John Wiley & Sons, 2008), 471–498.

incentive.”³⁷⁵ Additionally, if incentives are offered to participants, they are more likely to complete an online survey once started.³⁷⁶ Göritz found limiting payment of incentives to participants who complete a study has no impact on response quality or data compared with unconditional incentives.³⁷⁷

Interested participants completed a pre-questionnaire that identified the following: demographic characteristics; computer literacy level; previous experience with digital collections, archives, and social tagging; and knowledge level of the sample collection’s subject (see Appendix B for a copy of the assessment tool and Appendix C for a copy of the pre-questionnaire). Based on the questionnaire information, each participant was assigned to the novice or expert group unless the designated group reached its quota of thirty participants (in which case the participant will not be included in the study).

The knowledge level or expertise of a given participant was determined through completion of a brief ten-question multiple-choice assessment. The knowledge assessment focused on specific domain knowledge of the civil rights movement in Milwaukee, and was completed during the pre-questionnaire. The assessment questions were researched and developed by the author based on prior knowledge of the topic and the subject matters of the sample collection materials. Additionally, the assessment tool was reviewed by an independent researcher knowledgeable on the subject, and tested by several colleagues with a variety of knowledge levels.

³⁷⁵ Eleanor Singer and Richard A. Kulka, “Paying Respondents for Survey Participation,” in *Studies of Welfare Populations Data Collection and Research Issues*, ed. Michele Ver Ploeg, Robert A Moffitt, and Constance F Citro (Washington, DC: National Academy Press, 2002), 105–128.

³⁷⁶ Anja S. Göritz, “Incentives in Web Studies: Methodological Issues and Review,” *International Journal of Internet Science* 1, no. 1 (2006): 58–70.

³⁷⁷ Anja S. Göritz, “Contingent versus Unconditional Incentives in WWW-Studies,” *Advances in Methodology & Statistics* 2, no. 1 (2005): 1–14.

Based on the results, each participant's domain expertise was rated between 0 and 10 corresponding to the number of correct answers, and the participant was placed into one of three groups: novice (0-4, inclusively); intermediate (5-6, inclusively); or expert (7-10 inclusively). Participants falling within the intermediate range were excused from the study, thereby leaving a more polarized differential between study participants' knowledge levels. Through dismissing intermediate users, the dissertation avoids drawing conclusions from minuscule differences between those scoring a 4 and 5.

Participants provided demographic information by indicating inclusion within specified groupings in the areas of age, gender, level of education, and race. Participants also self-assessed their computer literacy level, experience with digital collections, archives, and experience with social tagging using a visual analog scale (VAS). According to Hasson and Arnetz, using a VAS for a single item can avoid the end-aversion bias of Likert scales where participants are less inclined to respond with either extreme.³⁷⁸ Hasson and Arnetz also found VAS more accurately identified participants' self-assessment of health than a Likert scale.³⁷⁹

3.3 Participant Demographics

The study's participants provided demographic information during a pre-questionnaire. Additionally, each of the participants answered a ten-multiple-choice question assessment of their prior Milwaukee Civil Rights movement knowledge. The assessment score divided participants into three groupings: experts (7-10, inclusively);

³⁷⁸ Dan Hasson and Bengt B. Arnetz, "Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements," *Global Journal of Health Education and Promotion* 8, no. 1 (2005): 178–192. See also Robert F. DeVellis, *Scale Development: Theory and Applications* (Newbury Park, CA: Sage, 1991).

³⁷⁹ Ibid.

intermediates (5-6, inclusively); and novices (0-4, inclusively). The intermediate participants did not continue with the study, and both the expert and novice groups reached the required thirty participants. The expert group had a mean score of 7.57 (n=30) with the novice group providing a mean of 2.77 (n=30).

The dissertation's six participants ranged in age from 18 to 63 with a mean age of 31.73, the median age of 28.5, and mode of 24 (n=60). The mean age of expert participants (\bar{x} = 35.1, n=30) skewed higher than novices (\bar{x} = 28.37, n=30). The majority of all participants were female, with similar gender divisions for both expert and novice groupings (see Table 3.3). Most participants came from either Wisconsin or Illinois (48.3%), although 21 twenty-one states and the District of Columbia are represented in the study (see Figure 3.1). The domain of the sample collection and the directed recruitment materials account for the high degree of response from Wisconsin or Illinois.

Table 3.3 Gender and Racial Characteristics of Participants

Demographic characteristic		Combined		Expert		Novice	
		No.	%	No.	%	No.	%
Gender	Male	14	23.3	7	23.3	7	23.3
	Female	45	75.0	22	73.3	23	76.7
	Other	1	1.7	1	3.3	0	0
Race*	White	36	60	18	60	18	60
	White, Black, & American Indian	1	1.7	1	3.3	0	0
	White & American Indian	2	3.3	0	0	2	6.7
	White & Other	1	1.7	1	3.3	0	0
	Black	7	11.7	6	20	1	3.3
	Hispanic/Latino	10	16.7	3	10	7	23.3

69.1%/30.9% racial division between white and non-white groupings, whereas the participants comprise a 66.7%/33.3% racial division.

The domain groupings create some interesting racial trends and divisions. Although those identifying only as white were equally distributed between experts and novices (eighteen per group), participants selecting only black primarily tested into the expert category (six experts and one novice). The disparity increases to seven if those participants who partially identified as black are included. The civil rights movement focus of the sample collection is likely associated with the high level of domain expertise among black participants since the assessment questions (and collection) focus on their ethnic group's history and culture. Additionally, the participants associated with the remaining non-white groups whose history is not specifically represented in the sample collection divide in the opposite direction with four experts and eleven novices.

The pre-questionnaire asked participants for their religious affiliation, including an option for not stating a preference (see Table 3.4). The majority of participants in both expert and novice groups identified as Christian (50%), with further divisions into Protestant (23.3%), Catholic (18.3%), and Evangelical (8.3%). Participants also identified highly with Atheism or Agnosticism (21.7%). Overall, the distribution of religious affiliation was relatively balanced between the expert and novice groups.

Table 3.4 Religious Affiliation of Participants

Religious	Combined		Expert		Novice	
Affiliation	No.	%	No.	%	No.	%
Prefer not to	12	20.0%	8	26.7%	4	13.3%

state						
Protestant	14	23.3%	9	30.0%	5	16.7%
Catholic	11	18.3%	3	10.0%	8	26.7%
Evangelical	5	8.3%	4	13.3%	1	3.3%
Jewish	2	3.3%	0	0.0%	2	6.7%
Muslim	0	0.0%	0	0.0%	0	0.0%
Hindu	1	1.7%	0	0.0%	1	3.3%
Buddhist	1	1.7%	0	0.0%	1	3.3%
Atheist/Agnostic	13	21.7%	5	16.7%	8	26.7%
Animistic	1	1.7%	1	3.3%	0	0.0%

Similar to the religious affiliation, the participants' reported highest completed education level is equally balanced between the expert and novice groupings (see Table 3.5). The majority of participants (58.3% of the combined totals) reported having completed some college or a bachelor's degree, while 38.3% of participants had completed postgraduate programs.

Table 3.5 Participants' Highest Completed Education Level

Completed	Combined		Expert		Novice	
Education Level	No.	%	No.	%	No.	%
High school or equivalent	3	5.0	2	6.7	1	3.3
Vocational	1	1.7	0	0	1	3.3
Some college	14	23.3	7	23.3	7	23.3
Bachelor's degree	19	31.7	9	30.0	10	33.3

Master's degree	16	26.7	9	30.0	7	23.3
Doctoral degree	4	6.7	1	3.3	3	10.0
Professional degree	3	5.0	2	6.7	1	3.3

The participants indicated their previous use of digital collections, archives, social tagging, knowledge of social tagging, and computer experience through a self-assessment on a visual analog scale (VAS) of zero to 100 during the pre-questionnaire. Table 3.6 reports the median and modes of the VAS scores for experts, novices, and the combination of both groups. Individual Mann-Whitney U tests were run to determine if there were differences in participants' self-assessed areas (prior use of digital collections, archives, social tagging, knowledge of social tagging, and computer experience) between expert and novices. For all five areas, the distribution of the area's levels for experts and novices were not similar, as assessed by visual inspection.

Prior use of digital collections' VAS scores for experts (mean rank = 30.42) and novices (mean rank = 30.58) were not statistically different, $U = 452.5$, $z = 0.037$, $p = 0.971$. Participants' prior use of archives VAS scores for experts (mean rank = 32.92) and novices (mean rank = 28.08) were not statistically different, $U = 377.5$, $z = 1.073$, $p = 0.283$. The participants' prior knowledge of social tagging VAS scores for experts (mean rank = 32.90) and novices (mean rank = 28.10) were not statistically different, $U = 378$, $z = 1.065$, $p = 0.287$; nor were the prior use of social tagging VAS scores for experts (mean rank = 32.43) and novices (mean rank = 28.57) statistically different, $U = 392$, $z = 0.86$, $p = 0.390$. Finally, computer experience level VAS scores for experts (mean rank = 28.52)

and novices (mean rank = 32.48) were not statistically different, $U = 509.5$, $z = 0.881$, $p = 0.378$.

The previously reported statistics and demographic information indicate a homogeneous composition of the dissertation participants. The following sections discuss the coding scheme, tag analysis, comparison with metadata and query log, and post-questionnaire data.

Table 3.6 Average VAS Scores from Pre-Questionnaire

	Computer Experience		Use of Digital Collections		Use of Archives	
	Median	Mean	Median	Mean	Median	Mean
Expert (n=30)	82.00	77.27	62.50	55.77	69.00	58.53
Novice (n=30)	85.00	83.07	61.00	55.07	54.50	49.67
Combined (n=60)	84.50	80.17	62.50	55.42	58.00	54.10
	Know. of Social Tagging		Use of Social Tagging			
	Median	Mean	Median	Mean		
Expert (n=30)	66.00	59.83	56.00	46.20		
Novice (n=30)	48.00	48.03	36.00	39.43		
Combined (n=60)	65.00	53.93	43.50	42.82		

3.4 Data Collection Methods and Procedures

Participant data collection during the study occurred in three phases: participant pre-questionnaire, tag generation, and participant post-questionnaire. Table 3.1 provides a breakdown of data collection and analysis methods by research question. Following pre-questionnaire completion and assignment to the expert or novice group, each participant viewed a brief video tutorial on how to submit tags within the CONTENTdm

environment. Upon completion of the video, further instructions directed participants to the sample collection on CONTENTdm.

Participants in both groups viewed and interacted with CONTENTdm in near-real-world conditions. Each group interacted with a duplicate of the sample collection in separate instances, and the initial users for each group did not see tags within the collection; however, subsequent participants viewed the tags added by previous users, thereby maintaining the look and feel of a regular digital collection. This helped simulate the normal generation of tags within collections. Each participant moved through each of the two sample sub-collections (documents and photographs) individually with the ability to move between records within the sub-collection.

Participants were randomly divided within their overall grouping into two subgroupings (expert 1, expert 2, novice 1, and novice 2). The use of random assignment and presenting the sample sub-collections in a different order normalized the resulting data and removed any influence of presentation order. The expert 1 and novice 1 subgroups first used and tagged the sample documents while the expert 2 and novice 2 subgroups initially tagged and used the sample photographs. Both subgroups from each domain group (expert, novice) viewed and tagged the same sample collection, with expert 1 and expert 2 tagging the expert sample collection and novice 1 and novice 2 tagging the novice sample collection.

Participants were required to submit at least one tag per item, but no limit was placed on the number of tags each participant could create. Participants could also submit duplicate tags if they agreed with a tag already provided by another user. This process

allowed the participant to virtually “approve” or “thumbs up” previous submissions. The required instructional video also directed participants only to provide English-language tags. This limitation was purely for analytical reasons, since non-English tags would be difficult to categorize beyond identification as non-English. Participants were not time-limited during the tagging exercise; however, participants spent an estimated 1-1.5 minutes per item for a total of 1-1.5 hours for the tagging activity.

Following the tagging exercise, participants completed a post-questionnaire containing a combination of structured and open-ended questions (see Appendix D). These questions focused on the participants’ tagging experiences and participants’ considerations during tags creation. Participants initially indicated their responses on 5-point Likert scales. Upon completion of the structure questions, participants could provide additional information for each category as prompted by the series of open-ended questions.

Participants viewed minimally processed metadata with the sample collection that was extracted from the existing *March on Milwaukee* digital collection. Additional metadata, that is the metadata not included within the minimally processed sample collection version, was extracted from the *March on Milwaukee* digital collection for all thirty items used in the dissertation project. The additional metadata referred to as “unselected metadata,” was aggregated into two lists (photographs and documents) and used for comparison with the generated tags during the evaluation.

One of the benefits of using a sample collection from an existing digital collection, in addition to the metadata extraction, is the ability to gather and analyze the searching

behavior of real users interacting with the collection. The digital librarian at the University of Wisconsin-Milwaukee Libraries, through her technical support office, shared the daily server logs for the entire Digital Collections at the UWM Libraries' CONTENTdm site for the month of January 2014. The server logs contained all websites visited for all digital collections, including user queries with query terms within the URLs. Individual URLs were extracted from each daily log and aggregated into a single list. Further parsing of the original list created two interrelated lists of user query terms through extracting the specific search terms from the collected URLs. One list included searches across all of the digital collections with 59,325 unique query terms. The second list focused on searches of the *March on Milwaukee* collection and included 1,609 unique query terms. Both lists were used for comparison with participants' tags during the data evaluation process for research question 3.

3.5 Pilot Study

A brief pilot study was conducted upon completion of building the data gathering devices (pre-questionnaire, sample collection, post-questionnaire), thereby verifying the usability of the tools themselves. Four participants were asked to walk through and complete the various stages of the study, and were informally interviewed afterward regarding any issues and/or suggested changes to the mechanisms. Minor alterations related to survey flow and directions were implemented following the pilot study. The four participants involved with the pilot study did not participate in the full study.

3.6 Data Analysis

Just as the dissertation project data comes from a variety of sources (questionnaires, tagging, existing metadata, and server logs), so too must the data

analysis. Overall, the data analysis combines several approaches in both qualitative and quantitative methods, thereby alleviating the limitations of one method with the strengths of another. A portion of the data analysis for all three research questions relies on multiple statistical analyses, therefore requiring clear delineations of the variables investigated. The independent variable for all statistical analyses is prior domain knowledge as defined through participant membership in one of three independent groups: expert, intermediate, or novice. Since the intermediate group members were excused from full participation in the study, only two independent groups comprise the independent variable. Membership in each of the domain knowledge groups is based on participants' scoring during the pre-questionnaire assessment; however, the knowledge level (and independent variable) is considered nominal since the assessment scores are used only to determine group membership and not to differentiate knowledge levels between members of the same group. In order to best address the proper data analysis, each research question and its associated analysis methodology, including the dependent variables and statistical tests applied, are discussed separately below.

3.6.1 RQ1a— What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?

The qualitative tag analysis relies on grouping the tags into categories and subcategories. Although coding schemes exist from previous studies, such as the Library of Congress Flickr Project, this study developed a new coding scheme based on an open coding of the data. The application of open coding allows “the categories and names for

categories to flow from the data,” rather than forcing the data into structured silos.³⁸¹

According to Corbin and Strauss:

In open coding, event/action/interaction, and so forth, are compared against others for similarities and differences; they are also conceptually labeled. In this way, conceptually similar ones are grouped together to form categories and their subcategories...Open coding and its characteristics of making use of questioning and constant comparisons enable investigators to break through subjectivity and bias. Fracturing the data forces examination of preconceived notions and ideas by judging these against the data themselves. A researcher can inadvertently attempt to place data into a category where it does not analytically belong, but by means of making systematic comparisons, these errors will eventually be located and the concepts placed in appropriate classifications.³⁸²

Since the coding process requires a comprehensive view of emerging categories, the tags from both experts and novices were merged into one group for analysis. The subsequent analysis identified six major categories (replication of metadata, format focused, subject, content summary, context, emotion, and incorrect) with one category (subject) containing two subcategories (general and specific). Table 3.7 lists and provides a definition for each category and subcategory. Section 4.1.2 in Chapter 4 discusses each of the categories in further detail.

³⁸¹ Hsiu-Fang Hsieh and Sarah E. Shannon, “Three Approaches to Qualitative Content Analysis,” *Qualitative Health Research* 15, no. 9 (2005): 1279.

³⁸² Juliet M. Corbin and Anselm Strauss, “Grounded Theory Research: Procedures, Canons, and Evaluative Criteria,” *Qualitative Sociology* 13, no. 1 (1990): 12–13.

Table 3.7 Coding Scheme Categories & Definitions

Category	Definition	Examples
Replication of Metadata	Tag duplicated information already included within minimal metadata	Father Groppi, hate mail, criticism mail
Format Focused	Tag identified, described, or otherwise focused on the format of the item	typed letter, black and white, handwritten
Subject—General	Tag identified objects, places, or people in the photograph or letter with common nouns	boy, cops, flag, gas mask
Subject—Specific	Tag identified objects, places, people, or dates in the photograph or letter with proper nouns and provided more specific information	1967, Beatrice Waiss, Marquette University, NAACP Youth Council
Content Summary	Tag summarized the photographed scene or letter contents	commando meeting, detained priest, police brutality, religious objection
Context	Tag placed photograph or letter within a broader context rather than discussing or identifying content within photograph or letter	desegregation, liberation theology, nationalism, race and religion
Emotion	Tag reflected an emotional response to photograph or letter	hope, inspirational, shame
Incorrect	Tag provided incorrect information	riot, music, criticism

Following the creation of the coding scheme, each tag was placed into a discrete category or subcategory. Once placed into categories and subcategories, the tags were tallied on a variety of levels, including a pure count of tags generated, tags in each category and subcategory, and total reductions from the record tallies, in order to provide an overall breakdown of tags by category/subcategory, record type, and participant group. To verify the coding scheme, an independent domain expert coded a random sample of

369 tags out of 9,278 (95% confidence level and confidence interval of 5). An analysis of the expert's codes found that 352 codes matched the researcher's resulting with a strong inter-coder reliability of 0.954 based on Holsti's reliability formula of $\frac{2M}{N_1+N_2}$.³⁸³

Additionally, Cohen's κ was run to further test the reliability of the coding scheme on the sample of 369 tags. According to the analysis, there was a very high level of agreement between the author and the expert coder, $\kappa = .943$ (95% CI, .916 to .970), $p < .0005$.

Descriptive statistical analysis summarized the findings' central tendency and dispersion.³⁸⁴

Part of research question 1(a) tested the association between the independent variable and the number of tags generated (dependent variable) in total, for the photograph set apart, and for the document set alone. Since the dependent variable in this case was continuous, and the independent variable consisted of two categorical independent groups, independent-samples t-tests were run based on the following three hypotheses:

H₁: The number of tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

H₂: The number of photographic tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

³⁸³ Ole R. Holsti, *Content Analysis for the Social Sciences and Humanities* (Reading, MA: Addison-Wesley, 1969).

³⁸⁴ Gillian Byrne, "A Statistical Primer: Understanding Descriptive and Inferential Statistics," *Evidence Based Library and Information Practice* 2, no. 1 (2007): 32–47; Barbara M Wildemuth, "Descriptive Statistics," in *Applications of Social Research Methods to Questions in Information and Library Science*, ed. Barbara M. Wildemuth (Westport, CT: Libraries Unlimited, 2009), 338–347; Barbara M. Wildemuth, "Frequencies, Cross-Tabulation, and the Chi-Square Statistic," in *Applications of Social Research Methods to Questions in Information and Library Science*, ed. Barbara M. Wildemuth (Westport, CT: Libraries Unlimited, 2009), 348–360.

H₃: The number of document tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

A second portion of research question 1(a) explored a possible association between the independent variable and type or category of tag created (dependent). In this instance, the dependent variable was also nominal, requiring Chi-square tests for association. The data analysis used three Chi-square tests based on the following hypotheses:

H₄: The proportion of tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

H₅: The proportion of photographic tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

H₆: The proportion of document tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

4.6.2 RQ1b— Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?

While research question 1(a) focuses on identifying similarities and differences between expert and novice users' tags, research question 1(b) explores potential differences in the tagging experience and things considered during tag creation. Participants indicated their opinions regarding both the experience and tag creation considerations through structured question responses on 5-point Likert scales during the post-questionnaire. Participants also provided additional information through open-ended questions following each grouping of structure questions. The tagging experience group

included three structured questions while the considerations group included five structured questions.

Research question 1(b)'s analysis included both statistical tests and content analysis of the open-ended responses. The statistical analysis tested for any difference between expert and novice users' opinions on the factor-based aspects (dependent variables). In each case, the Likert scale responses were ordinal rather than continuous and required Mann-Whitney U tests rather than t-tests. The Mann-Whitney U tests were based on the following hypotheses:

H₇-H₉: Expert and novice users' opinions of the tagging experience are different for ease of tagging in general (H₇); difficulty in tagging documents compared to photographs (H₈); and difficulty in tagging photographs compared to documents (H₉).

H₁₀-H₁₄: Expert and novice users' opinions of the considerations for the creation of tags are different for how others would find the item (H₁₀); how the tagger (user) would find the item (H₁₁); the content of the tagged item (H₁₂); the format of the tagged item (H₁₃); and other users' tags (H₁₄).

3.6.3 RQ2a— In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive? RQ2b— Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?

The sample digital archives contain a subset of the original metadata in the existing *March on Milwaukee* digital collection. Addressing RQ2 required a comparison of the generated tags from both experts and novices with the unselected metadata from

the existing collection. A comparison group of unselected metadata was generated for each sample record group (document and photograph) including the fields from the following Dublin Core elements: title, creator, subject, description, date, format, identifier, and language. The unselected metadata lists were filtered through a stop list prior to additional analysis since several fields included several non-descriptive terms (such as articles). The comparison of unselected metadata and tags considered only exact matches rather than partial or matching word variations. The analysis generated descriptive statistics for each format grouping, highlighting the number and percent of matching terms, and the number and percent of new terms for both expert and novice groups.

Although the users' knowledge level was initially assessed during the pre-questionnaire, this information was used only to put the participants into categorical groupings, and not to differentiate knowledge levels within groupings during later analysis (e.g., participant one is more of an expert than participant two). Since the independent variables (user knowledge) are, therefore, categorical (or nominal) rather than quantitative, and a Chi-square test best fit the needs of the research question. A 2 x 2 table Chi-square test for association based on the numerical values (number matching and number not matching) tested the following hypothesis:

H₁₅: The proportion of tags matching unselected metadata is affected by the user's domain knowledge.

The researcher also calculated the Phi and Cramer's V to analyze the strength of any potential relationships between group type and the number of matching terms. The

strength of association test used will be Phi since the X^2 analysis was based on a 2 x 2 table.

3.6.4 RQ3a— In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive? RQ3b— Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?

The data analysis addressing RQ3 followed a similar process to that of RQ2. Rather than looking at format-based groupings, however, this analysis focused on the entire sample collection. The query terms from actual users were parsed out of the existing server-log data and used as a comparison group. Parsing of the server logs resulted in 59,325 unique query terms used to search across all collections hosted by UWM-DC. Further reduction by collection-specific searches found 1,609 unique query terms used to search the *March on Milwaukee* collection alone. A list of unique tag terms created by each domain group (expert, novice) and a third list with all unique tag terms created were compared to both query term lists. Additionally, the unique unselected metadata terms were also compared to the *March on Milwaukee* query term list. The comparisons considered only exact matches rather than partial or matching word variations. The analysis generated descriptive statistics highlighting the number and percent of matching terms, and the number and percent of non-matching terms for expert and novice tags, the combination of expert and novice tags, and the unselected metadata.

Research question 3(b) utilized Chi-square tests for association to explore potential relationships between the independent variable and the proportion of tags matching user query terms, the dependent variable. Similar to previous, Chi-square tests

were selected since the dependent variables were nominal; specifically, matching or not-matching being the dichotomous categories. This analyzed the following hypothesis:

H₁₄: The proportion of tag terms matching users' query log terms is affected by user's domain knowledge.

The researcher also calculated the Phi and Cramer's V to analyze the strength of any possible relationships between group type and the number of matching terms. The strength of association test used will be Phi since the X^2 analysis was based on a 2 x 2 table.

3.7 Validity, Reliability, and Generalizability

Ensuring the validity of findings offers the largest challenge of any experimental design. Although experimental designs provide the opportunity for high degrees of control, errors in planning may lead to questionable conclusions, thereby putting the entire process in jeopardy. Validity occurs both internally and externally, with internal validity concerned with authenticating the observed relationship between independent and dependent variables. Otherwise stated, are the data and its indications an accurate reflection of the experiment or did some outside force negatively influence the results? While internal validity explores the experimental conclusions themselves, external validity concerns the experiment's real-world application. In other words, can the findings be applied to other groups or generalized for the general population?³⁸⁵

³⁸⁵ Donald T Campbell and Julian C Stanley, *Experimental and Quasi-Experimental Design for Research* (Chicago: Rand-McNally, 1963); John W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed. (London: Sage, 2009); Diane Kelly, "Methods for Evaluating Interactive Information Retrieval Systems with Users," *Foundations and Trends in Information Retrieval* 3, no. 1–2 (2009): 1–224; Ronald R Powell and Lynn Silipigni Connaway, *Basic Research Methods for Librarians*, 4th ed. (Westport, Conn.: Libraries Unlimited, 2004); Steven M. Ross and Gary R. Morrison,

Regarding the dissertation project, the population sampling technic and overall quasi-experimental design reduces internal validity threats. The use of assessment for group assignment, and excluding participants in the intermediate group, for example, limit regression threats. Although pure random assignment to the groups was not possible, careful group membership selection with an eye toward homogeneity limits the effects of selection bias.³⁸⁶ Furthermore, the inclusion of participation incentives for completion of the study reduces the mortality threat.³⁸⁷ The dissertation project does not encounter treatment-based internal validity threats since the participants will not interact with each other.³⁸⁸

External validity concerns the generalizability of an experiment's findings. Strengthening external validity typically involves loosening experimental controls, which, therefore, decreases internal validity. The dissertation project design addresses several external validity threats. The online nature of both the population sampling and study limits the interaction of selection and treatment threat, common and dangerous threat among library and information science research.³⁸⁹ Unlike the convenient sampling techniques typically used, researchers using Internet-based experiments have a larger

"Handbook of Research for Educational Communications and Technology," 1996, <http://aect.org/edtech/ed1/>.

³⁸⁶ Creswell, *Research Design*; Carolyn Hank and Barbara M. Wildemuth, "Quasi-Experimental Studies," in *Applications of Social Research Methods to Questions in Information and Library Science*, ed. Barbara M.M Wildemuth (Westport, CT: Libraries Unlimited, 2009), 93–104; A.I. Piper, "Conducting Social Science Laboratory Experiments on the World Wide Web," *Library and Information Science Research* 20, no. 1 (1998): 5–21.

³⁸⁷ Hank and Wildemuth, "Quasi-Experimental Studies"; Kelly, "Methods for Evaluating Interactive Information Retrieval Systems with Users"; Barbara M. Wildemuth and Leo L. Cao, "Experimental Studies," in *Applications of Social Research Methods to Questions in Information and Library Science*, ed. Barbara M. Wildemuth (Westport, CT: Libraries Unlimited, 2009), 105–115.

³⁸⁸ Creswell, *Research Design*.

³⁸⁹ Ibid.

potential pool of participants representing a larger demographic slice of the population.³⁹⁰ Additionally, participants could complete all elements of the experiment from within their own home, thereby increasing their volunteering likelihood. The dissertation also discusses the interaction of setting and the treatment threat.³⁹¹ This threat restricts generalizability of experimental results based upon the experiment's setting. If conducted within a laboratory, for example, a study may not be generalizable to real-world settings. The dissertation attempts to simulate real-world settings whenever possible to address this threat.

³⁹⁰ Piper, "Conducting Social Science Laboratory Experiments on the World Wide Web."

³⁹¹ Creswell, *Research Design*.

CHAPTER FOUR: RESULTS

Following a two-month recruitment process, the researcher analyzed the data collected from the sixty participants. The participants generated 9,278 tags of which 1,463 were unique. Novices created more tags on average than experts, but experts created more unique tags than novices. An open-coding analysis developed a six-category coding scheme with one category containing two subcategories. In the aggregate, experts created more content-summary tags while novices created more emotion, format-focused, subject, and context tags. Both expert and novice participants' opinions regarding selected tagging factors found they enjoyed the tagging experience; however, they also indicated a desire for additional tagging system support. Additionally, they both considered rewards (non-monetary and/or monetary) a highly motivating reason for future tagging.

When compared with the unselected metadata, the generated tags mainly matched unselected metadata from the Dublin Core elements title, subject, and description. Additionally, document tags matched the unselected metadata more frequently than photograph tags in the title, subject and format Dublin Core fields. Although expert tags matched the unselected metadata more than novice tags, the combination of both expert and novice tags provided the highest proportion of matching terms. Expert tags matched the user query terms more often than novice tags, with the combination of both groups receiving the largest number of matching query terms.

This chapter details the dissertation study's data and resulting data analysis findings through discussing the following three research questions, their associated sub-questions, and hypotheses:

- Research Question 1(a): What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?
 - H₁: The number of tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₂: The number of photographic tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₃: The number of document tags generated in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₄: The proportion of tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₅: The proportion of photographic tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
 - H₆: The proportion of document tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.
- Research Question 1(b): Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?
 - H₇-H₉: Expert and novice users' opinions of the tagging experience are different for ease of tagging in general (H₇); difficulty in tagging

documents compared to photographs (H_8); and difficulty in tagging photographs compared to documents (H_9).

- H_{10} - H_{14} : Expert and novice users' opinions of the considerations for the creation of tags are different for how others would find the item (H_{10}); how the tagger (user) would find the item (H_{11}); the content of the tagged item (H_{12}); the format of the tagged item (H_{13}); and other users' tags (H_{14}).
- Research Question 2 (a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?
- Research Question 2 (b): Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?
 - H_{15} : The proportion of tags matching unselected metadata is affected by the user's domain knowledge.
- Research Question 3 (a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?
- Research Question 3 (b): Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?
 - H_{16} : The proportion of tag terms matching users' query log terms is affected by user's domain knowledge.

4.1 Research Question 1(a): What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?

The following section discusses the results of the study related to the scope of research questions 1(a) beginning with a comparison of the number of tags generated by expert and novice participants during the experiment. The second subsection provides a detailed description of the type and categories of tags created by both groups, providing general trends and characteristics of the tags. The final section highlights the specific similarities and differences between expert and novice tags.

4.1.1 Number of Tags Generated by Expert and Novice Participants

The study required each participant to create at least one tag per item for fifteen photographs and fifteen documents. Although the experts and novices interacted with separate identical versions of the sample collection, and, therefore could not see the tags generated by another domain group, they could view (and reuse) tags created within their own domain group. Combined, the participants generated a wide range of tags, from the required minimum thirty to one participant creating 1,031 tags. The novice participants generated more tags on average than the experts, with 57% of novices creating more than 115 total tags compared to 43% of experts. Table 4.1 presents the aggregate tag counts by format and users including the number of unique tags. Figures 4.1 and 4.2 chart the number of tags generated by each participant divided by format.

At first glance, novice users appear to generate a significantly higher number ($\bar{x} = 169.3$, $n = 30$) of tags than experts ($\bar{x} = 112.1$, $n = 30$); however, the tag generation of

three participants (two experts and one novice) skewed the overall data. E8, E26, and N28 each created over 500 total tags during the study and are considered outliers as confirmed by a box-plot analysis (see Figure 4.3). Removing these outliers reduces the gap between novices and experts from an average difference of 57.2 to 27.49. Due to these issues, the outliers were removed prior to subsequent statistical analysis.

Following the removal of outliers, an assessment by Shapiro-Wilk's test found the number of all tags created for each domain group was not normally distributed ($p < .05$). Further assessment by Shapiro-Wilk's tests found the number of photographic tags generated for each domain group was normally distributed ($p > .05$) while the number of document tags was not normally distributed ($p < 0.5$). Data are mean \pm standard deviation, unless otherwise stated. There were 28 expert and 29 novice participants. The novices produced more tags combined (139.59 ± 85.48) than experts (112.07 ± 62). Novices made more photographic tags (53.97 ± 31.53) than experts (47.43 ± 26.67). Finally, novices also generated more document tags (85.62 ± 60.63) than experts (64.64 ± 39.62).

Independent-samples t-tests were run to determine if there were differences in the three tag categories (all tags, photographic tags, and document tags) between experts and novices. The t-tests used the following hypotheses:

H₁: The number of tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

H₂: The number of photographic tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

H₃: The number of document tags generated in a minimally processed digital archive is affected by a user's domain knowledge.

There was homogeneity of variances for expert and novices, as assessed by Levene's test for equality of variances, for all tags ($p = .165$), photographic tags ($p = .185$), and document tags ($p = .376$). There was not a statistically significant difference in the mean number of combined tags generated between experts and novices, although novices averaged more than experts, 27.51 (95% CI, -67 to 12), $t(55) = -1.387$, $p = .171$.

Analyzing the document tags also found there was not a statistically significant difference between experts and novices, with novices averaging more than experts, 20.98 (95% CI, -48.3 to 6.3), $t(55) = -1.540$, $p = .129$. Finally, the analysis of photographic tags found there was not a statistically significant difference in the mean number of tags generated between experts and novices, with novices again averaging more than experts, 6.5 (95% CI, -22 to 9), $t(55) = -0.844$, $p = .403$.

Overall, while novice participants produced more tags than expert participants, independent-samples t-tests with and without the outlier users indicated the differences were not statistically significant. The lack of statistical significance indicates domain knowledge does not affect the number of tags generate. Both groups averaged above the minimum of 30 tags demonstrating indicating most participants did not merely consider the minimum requirements for the study. Additionally, both experts and novices produced more tags for the documents than the photographs, most likely due to the ease of adding words appearing within the documents over identifying tags associated with images. Finally, expert participants created more unique tags than the novices for both photographs and documents.

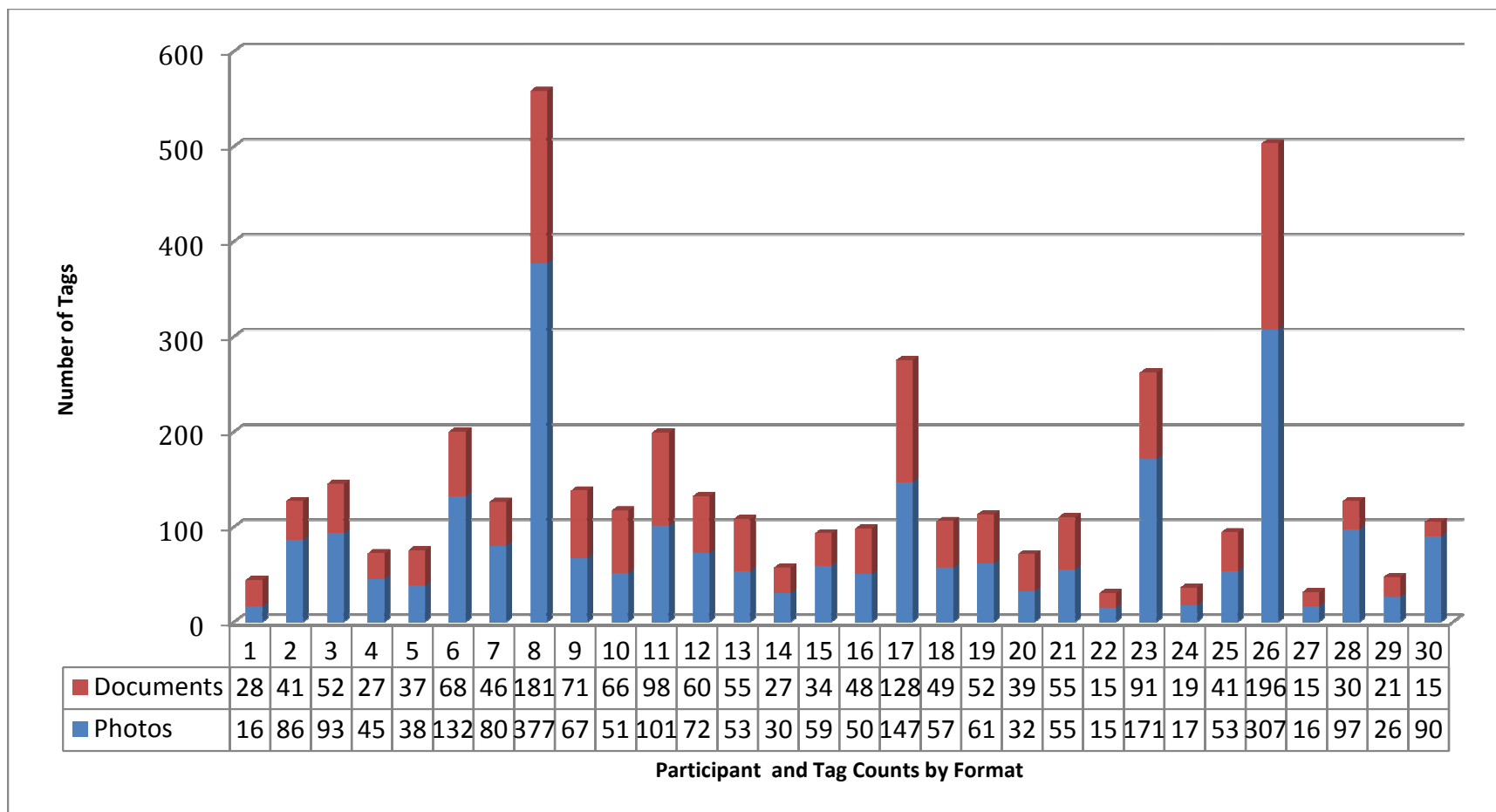


Figure 4.1 Expert Tag Counts by Format

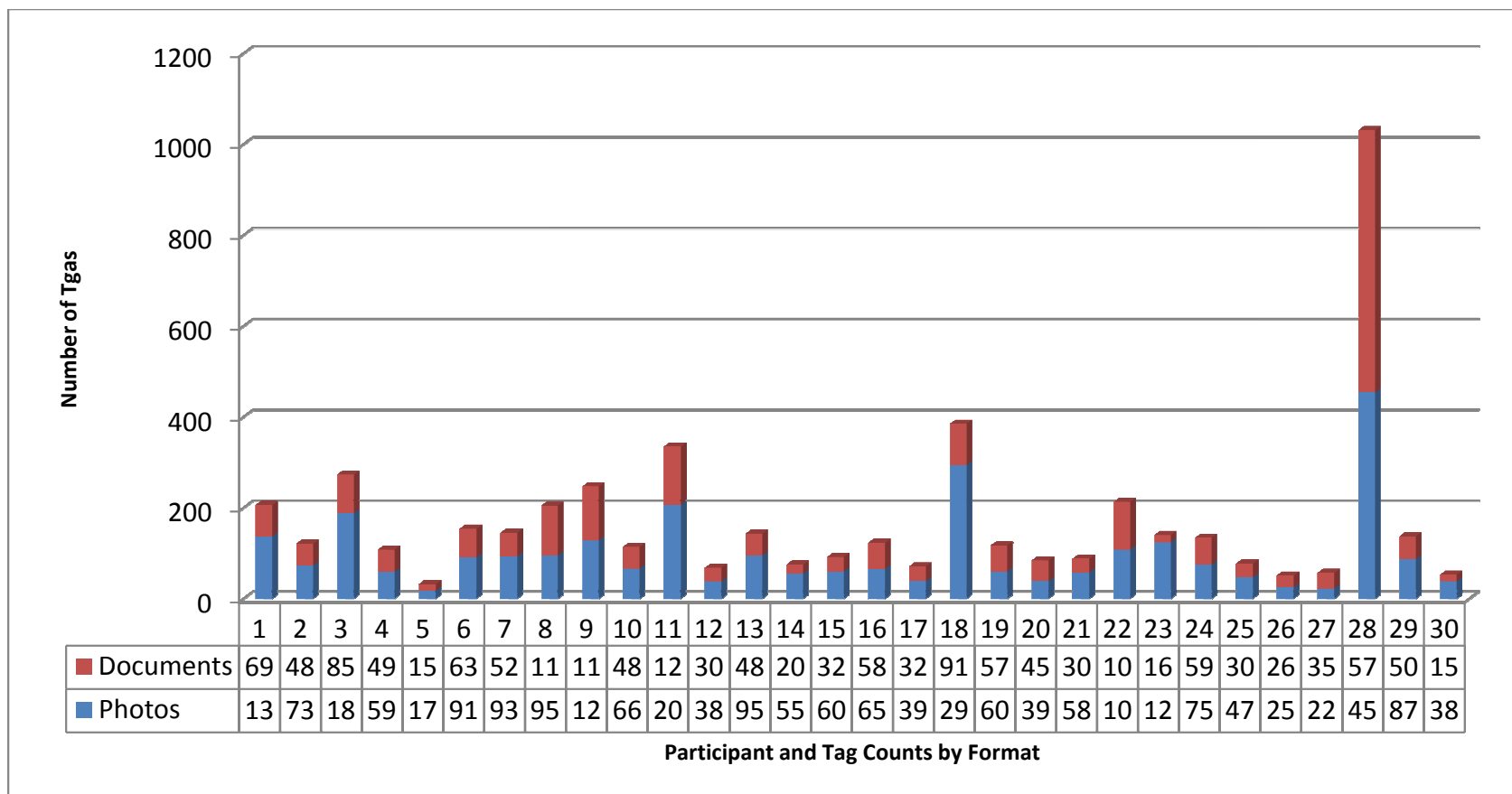


Figure 4.2 Novice Tag Counts by Format

Table 4.1 Aggregate Tag Counts by Users and Format

	Users	Total	Unique	Min	Max	Mean	Median	Mean w/o Outliers*
Photographs	Expert	1705	396	15	196	56.83	47	47.43
	Novice	2142	293	15	577	71.4	48.5	53.97
	Expert & Novice	3847	573	15	577	64.12	48	50.75
Documents	Expert	2494	685	15	377	83.13	58	64.64
	Novice	2937	579	17	454	97.9	69.5	85.62
	Expert & Novice	5431	995	15	454	90.52	63	73.32
Combined	Expert	4199	1020	30	558	139.97	109	112.1
	Novice	5079	805	32	1031	169.3	122	139.59
	Expert & Novice	9278	1463	30	1031	154.63	115.5	126.1

* Recalculated means without three outlier participants: E8, E26, and N28; For recalculation, Experts, n = 28, Novices, n = 29

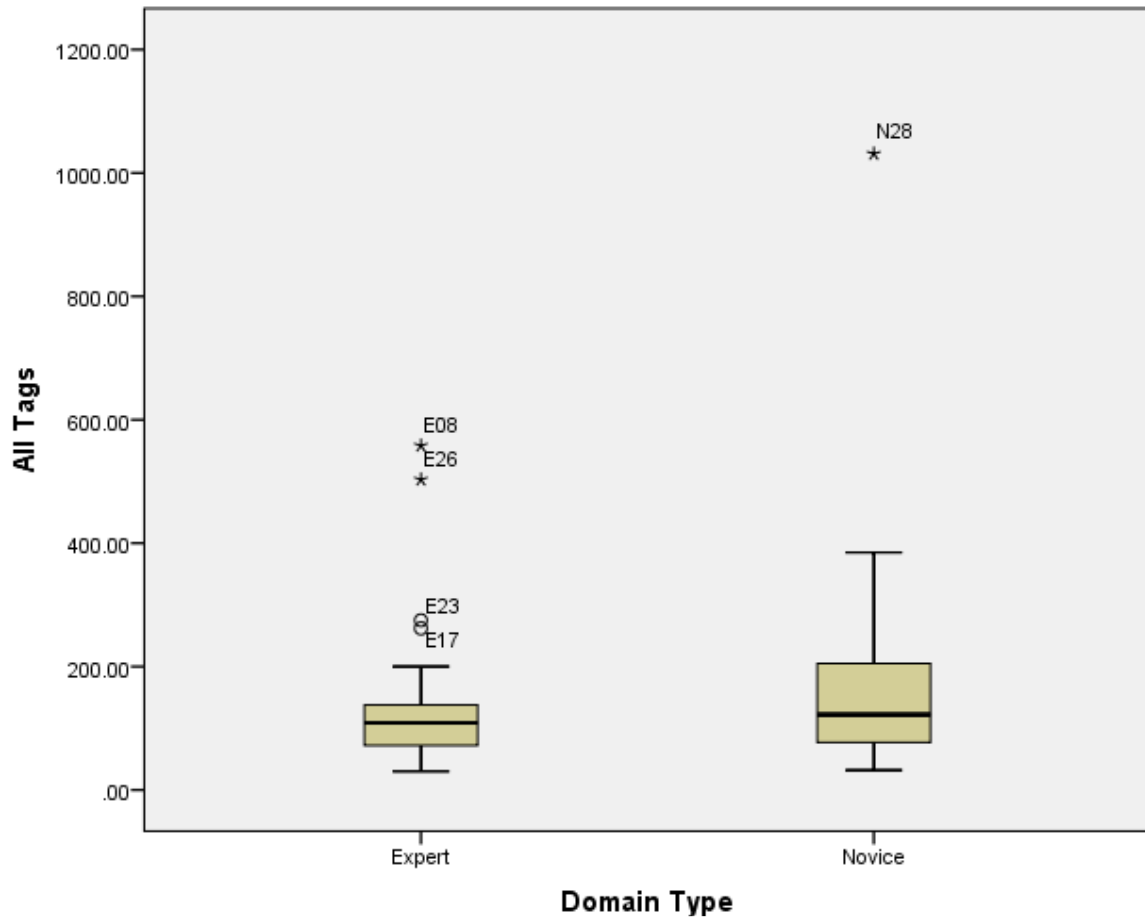



Figure 4.3 Box-plot Analysis of Number of Tags Generated by Participants

4.1.2 Types of Tags Generated by Expert and Novice Participants

The initial coding analysis of the 9,278 tags identified six major categories and two subcategories. An additional major category was added to the six following the inter-coder reliability testing phase. The final coding scheme, therefore, includes seven major categories: replication of metadata, format focused, subject, content summary, context, emotional, and incorrect. The category of subject is further broken down into two subcategories: general and specific. The following section describes the various categories and provides examples for both documents and photographs (see Tables 4.2 & 4.3 for a summary and examples of the coding scheme).

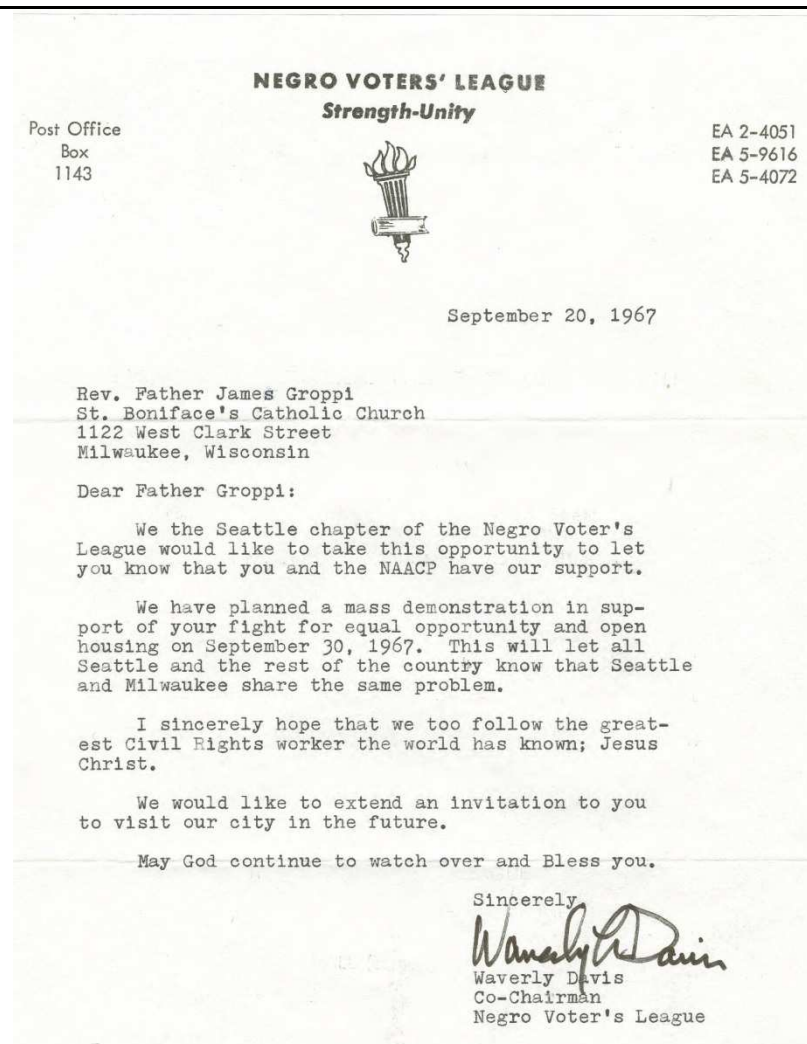
Table 4.2 Examples of Photograph Tags by Category

	Category	Examples
	Replication of Metadata	Groppi, Father Groppi, photograph
	Format Focused	black and white, black-and-white photography
	Subject—General	big man, police, riot gear, wagon
	Subject—Specific	Wagon 722, 1967, Milwaukee Police
	Content Summary	arrested, detained priest, inside police vehicle
	Context	Catholic social action, civil rights movement, race
	Emotion	unjust, acceptance
	Incorrect	courtroom*
*Note: The specific example provided did not occur within study data, but represents the type of tag typically found in the incorrect category.		

Wisconsin Historical Society, WHi-26541

Table 4.3 Examples of Document Tags by Category

Category	Examples
Replication of Metadata	Father Groppi, support mail
Format Focused	typed, typewritten
Subject—General	demonstration, voter, housing
Subject—Specific	1967, St. Boniface, NAACP
Content Summary	religious support, Seattle comparison
Context	race and religion, religious activism
Emotion	inspirational
Incorrect	hate mail, riot



Wisconsin Historical Society, WHi-111269

The first major category, replication of metadata, included tags which duplicated information already presented to the user in the minimal metadata for each item. The minimal metadata included information from the following fields: Title,³⁹² Part of Collection, Creator, Type (DCMI), Original Collection, Original Item Location, Original Item Location, Original Item Type, Finding Aid, Repository, Digital Publisher, Date Digitized, Digital Format, Digital Collection, and Rights. These tags reinforce the previous findings of Jeong’s YouTube analysis, although at much lower rates (as will be discussed later).³⁹³



Figure 4.4 Tag Cloud of all Replication of Metadata Tags

Table 4.4 Most Frequent Replication of Metadata Tags

Tag	Frequency	Percent
Groppi	440	25.7%
Father Groppi	353	20.6%
letter	168	9.8%
James Groppi	146	8.5%
Rev. Groppi	86	5.0%
hate mail	63	3.7%
support	60	3.5%
photograph	56	3.3%
Rev. James Groppi	43	2.5%
criticism	42	2.5%

³⁹² The Title field did not include the official, item-level description title of the object. Rather, a more generic title was used, such as Photograph 1.

³⁹³ Jeong, “Does Tagging Really Work?” and Jeong, “Is Tagging Effective?—Overlapping Ratios with Other Metadata Fields.”

Figure 4.4 illustrates a word cloud for all replication tags and Table 4.4 lists the most frequent tags. Combined, the replication tags represented 18.47% of all tags created. Although several different tags fit this grouping, the most commonly applied was Fr. Groppi or some variation thereof. The tags referencing Fr. Groppi made up 66.6% of all replication tags. Participants also tended to use the generic title of the item as a tag (e.g., “photograph” for Photograph 1, “support letter” for Support Letter 1, etc.); this occurred in 29.4% of replication tags. Although there was a difference in replication tag use frequency between experts and novices (discussed later), the general nature of the use and the tags themselves did not differ.

The second major category included tags focused on the format of the items themselves (see Figure 4.5 and Table 4.5). The third least used category at 1.33% of all tags, format tags highlighted the nature of the tagged items. Participants applied two different tags, “black and white” and “black-and-white photography,” for the photographic items. Additionally, only novices used format tags within the photographs. Within the document set, the format category mainly identified if the document was typed or handwritten. A few additional tags further delineated the handwriting as “illegible.”



Figure 4.5 Tag Cloud of all Format-Focused Tags

Table 4.6 Most Frequent Subject—General Tags

Tag	Frequency	Percent
priest	136	5.7%
police	106	4.5%
Father	63	2.6%
white	57	2.4%
black	46	1.9%
youth	43	1.8%
Reverend	42	1.8%
flag	36	1.5%
riot	34	1.4%
children	32	1.3%

The combined tag analysis found 25.64% as subject—general and 23.85% as subject—specific. Although the combination of photograph and document tags found a close division between general and specific subject, separating the formats revealed an intriguing difference. The photograph tags’ general/specific gap is 13.1 percentage points in favor of general (25.24%/12.14%) whereas the document tags’ general/specific gap is 6.22 percentage points in favor of specific (25.93%/32.15%). The formats themselves explain the difference since the documents provided participants directly with proper nouns to use as tags

Table 4.7 Most Frequent Subject-Specific Tags

Tag	Frequency	Percent
1967	288	13.0%
Catholic	196	8.9%
Milwaukee	194	8.8%
NAACP	93	4.2%
NAACP Youth Council	58	2.6%
commandos	52	2.3%
south side	50	2.3%
August	41	1.9%
God	38	1.7%
1966	35	1.6%

tags. Similar to the subject tags, the nature of the formats reveal the format disparity since the photographs required more interpretation, they produced a higher percent of the content-summary tags (1,051 out of 1,514 tags or 69.4%). The photograph content-summary tags often incorporated the entire idea of an image, whereas the document content summaries sometimes focused on one paragraph rather than the entire document.

Tags in the fifth major category contextualized the object (see Figure 4.9 and Table 4.9) and represent 13% of all tags. Often these tags focused on the Civil Rights Movement or a theme within the movement, such as race, segregation, non-violence, solidarity, or religion. Although these terms appear as tags within other categories, it is their use in relation to the specific item tagged that placed them into separate categories. Participants applied the tag, “black power,” for example, to Letter 2 in Criticism Mail. Since the phrase “black power” appears within the letter (see Figure 4.10), these tags are identification—general. Participants used the same tag for Photograph 11 (see Figure 4.11), and since “black power” does not specifically appear within the image, and functions more as a contextualization of the image, this occurrence of the tag fits better in the context category.

Table 4.9 Most Frequent Context Tags

Tag	Frequency	Percent
civil rights	219	18.2%
Civil Rights Movement	138	11.4%
Milwaukee	83	6.9%
race	73	6.1%
racism	60	5.0%
segregation	47	3.9%
Catholic	43	3.6%
religion	43	3.6%
Catholicism	29	2.4%
bussing	24	2.0%

8-30-67

Dear Father? Groppi,
 Don't crush this letter - Read it!
 you are a trouble maker and I'm
 sure you know that, you, a priest,
 I am a Catholic and ashamed of
 writing to you for something like
 this.

I must say you are very coy.
 Looking endlessly for some perfectly
 forest, yet inxigative idea.

Come to the South side, very right-
 eously (granted) asking the white
 with your yelling "black power" etc,
 of which they have become quite
 acquainted ever since they heard
 of you and your "commandos." And
 then when you did your best to instig-
 ate the crowd - who, I must say, had
 no business being there (but were
 exkindled with a feeling of hate - which
you - Father? Groppi helped to install)

Figure 4.10 Criticism Mail Letter 2, Wisconsin Historical Society, WHi-111271



Figure 4.11 Photograph 11, Wisconsin Historical Society, WHi-53596



Figure 4.12 Tag Cloud of all Emotion Tags

Table 4.10 Most Frequent Emotion Tags

Tag	Frequency	Percent
anger	14	13.7%
angry	11	10.8%
shame	10	9.8%
hate	10	9.8%
Hope	8	7.8%
happy	8	7.8%
Joy	3	2.9%
freedom	3	2.9%
ashamed	3	2.9%



Figure 4.13 Tag Cloud of all Incorrect Tags

Table 4.11 Most Frequent Incorrect Tags

Tag	Frequency	Percent
riot	16	59.3%
catholic hate	3	11.1%
criticism	3	11.1%
hate mail	3	11.1%
music	2	7.4%

The penultimate major category included tags containing an emotional response to one of the objects (see Figure 4.12). The emotion tags occurred in small numbers (1.1% of all tags) and slightly more often in photographs than documents (1.4% of photograph tags, 0.88% of document tags).

The last major category was reserved for incorrect tags (see Figure 4.13). The original coding scheme did not include the last category; however, after discussion with the outside coder used for inter-coder reliability, and reconsideration of previous research, the category appeared necessary. Although the author occasionally did not fully agree with the participants' interpretations of the photographs or documents, tags that merely gave a different interpretation were not placed into the incorrect category. The tag analysis only put tags without any association with the photograph or document into the incorrect category.

Surprisingly, only 27 (out of 9,278) or 0.29% of all tags were identified as being incorrect, and the vast majority of these came from two participants (see Figure 4.13). Participant E26 provided 14 incorrect tags (51.9%) and Participant N23 added 9 incorrect tags (33.3%); combined the two participants account for 85.2% of all incorrect tags. Each of the two participants gave different patterns of incorrect tags. Participant E26 produced the highest number of tags (503) but used the tag "riot" for 14 of his/her incorrect tags. Alternatively, Participant N23 produced a relatively average number of tags (140) and used three different tags incorrectly (catholic hate, criticism, and hate mail) all within the support mail letters.

Table 4.12 provides the categorical disbursement for photograph, document, and all tags; Figure 4.14 further illustrates each grouping. As an aggregate, the top three tag categories were: Subject—General (25.64%), Subject—Specific (23.85%), and Replication of Metadata (18.47%).

When analyzed by format, the top categories both differ from each other and the aggregate level. Photographs primarily fell into Content Summary (27.32%), Subject—General (25.24%), and Context (16.35%), while documents more closely aligned with the aggregate: Subject—Specific (32.15%), Subject—General (25.93%), and Replication of Metadata (20.95%). The close relationship between the aggregate and document-specific categorizations is primarily caused by the higher number of document tags (compared to photograph tags) influencing the aggregate level.

Table 4.12 Tag Counts and Percentages by Category and Format

	Photographs (n = 3847)		Documents (n = 5431)		Combined (n = 9278)	
	No.	%	No.	%	No.	%
Replication of Metadata	576	14.97%	1138	20.95%	1714	18.47%
Format Focused	89	2.31%	34	0.63%	123	1.33%
Subject—General	971	25.24%	1408	25.93%	2379	25.64%
Subject—Specific	467	12.14%	1746	32.15%	2213	23.85%
Content Summary	1051	27.32%	463	8.53%	1514	16.32%
Context	629	16.35%	577	10.62%	1206	13.00%
Emotion	54	1.40%	48	0.88%	102	1.10%
Incorrect	10	0.26%	17	0.31%	27	0.29%

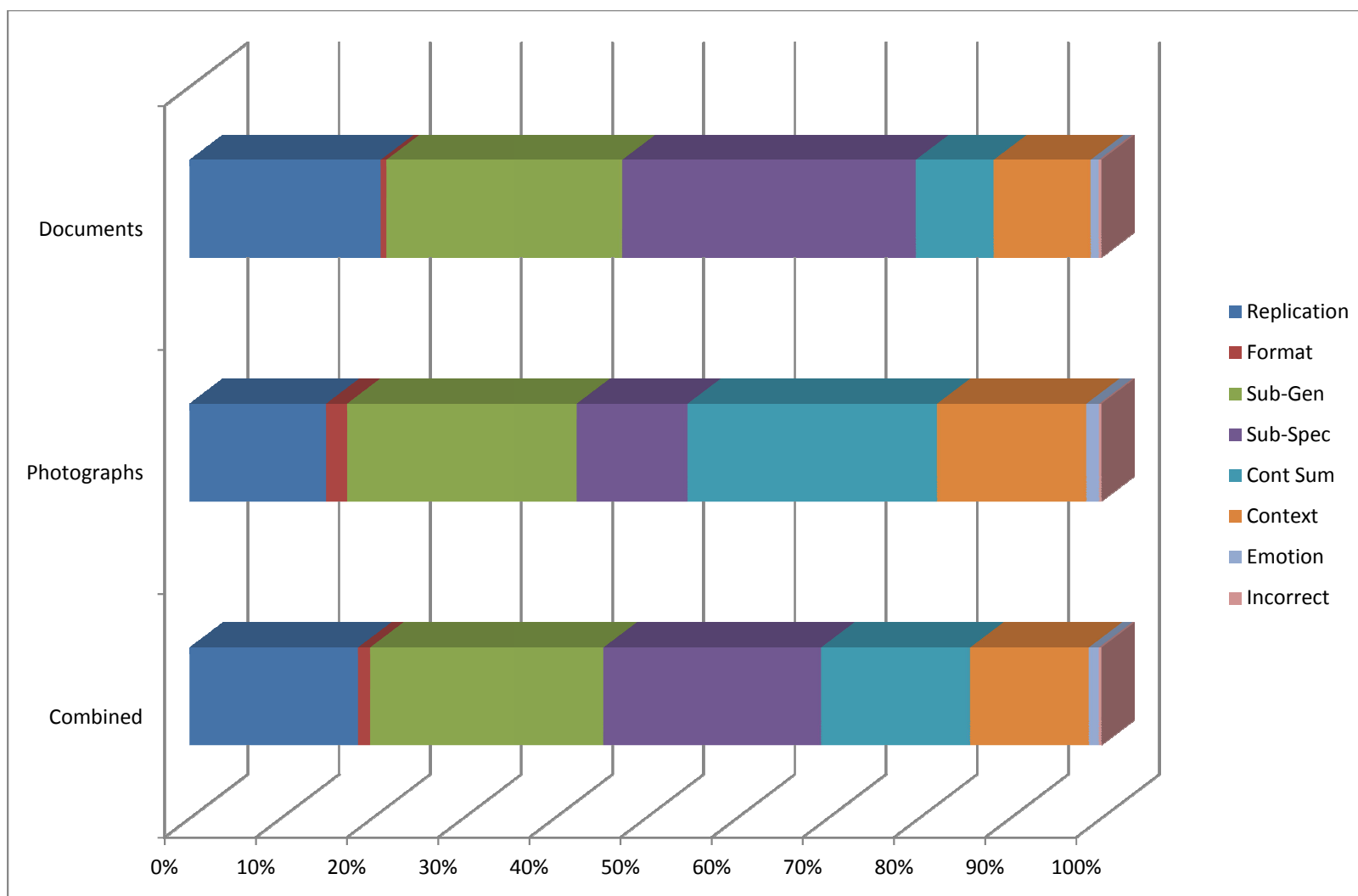


Figure 4.14 Comparison Expert & Novice Tag Categories Percentage by Format

4.1.3 Similarities and Differences of Expert and Novice Participants' Tags

While the previous section noted some differences between experts and novices, this section focuses on a direct comparison of the two groups' tags following the coding analysis. Comparing expert and novice tags for photographs and documents reveals some initial similarities and differences (see Table 4.13 and Figure 4.15). The main similarities with both expert and novice tags focus on potential issues with user-generated tags. Both domain groups replicated the minimally processed metadata at nearly identical rates (18.69% and 18.29%). At almost a fifth of all created tags, these tags did not contribute any new access points or description of the tagged objects. Both experts and novices rarely created incorrect tags, the implications of which are further discussed in the following chapter. Novices provided twice the amount of emotion tags and more than double the number of format-focused tags. Novices used slightly more context, subject—general, and subject—specific tags. Experts, on the other hand, created more content-summary tags.

Table 4.13 Number and Percent of All Expert and Novice Tags by Category

	Experts (n = 4199)		Novices (n = 5079)	
	No.	%	No.	%
Replication of Metadata	785	18.69%	929	18.29%
Format Focused	26	0.62%	97	1.91%
Subject—General	1022	24.34%	1357	26.72%
Subject—Specific	997	23.74%	1216	23.94%
Content Summary	791	18.84%	723	14.24%
Context	532	12.67%	674	13.27%
Emotion	30	0.71%	72	1.42%
Incorrect	16	0.38%	11	0.22%

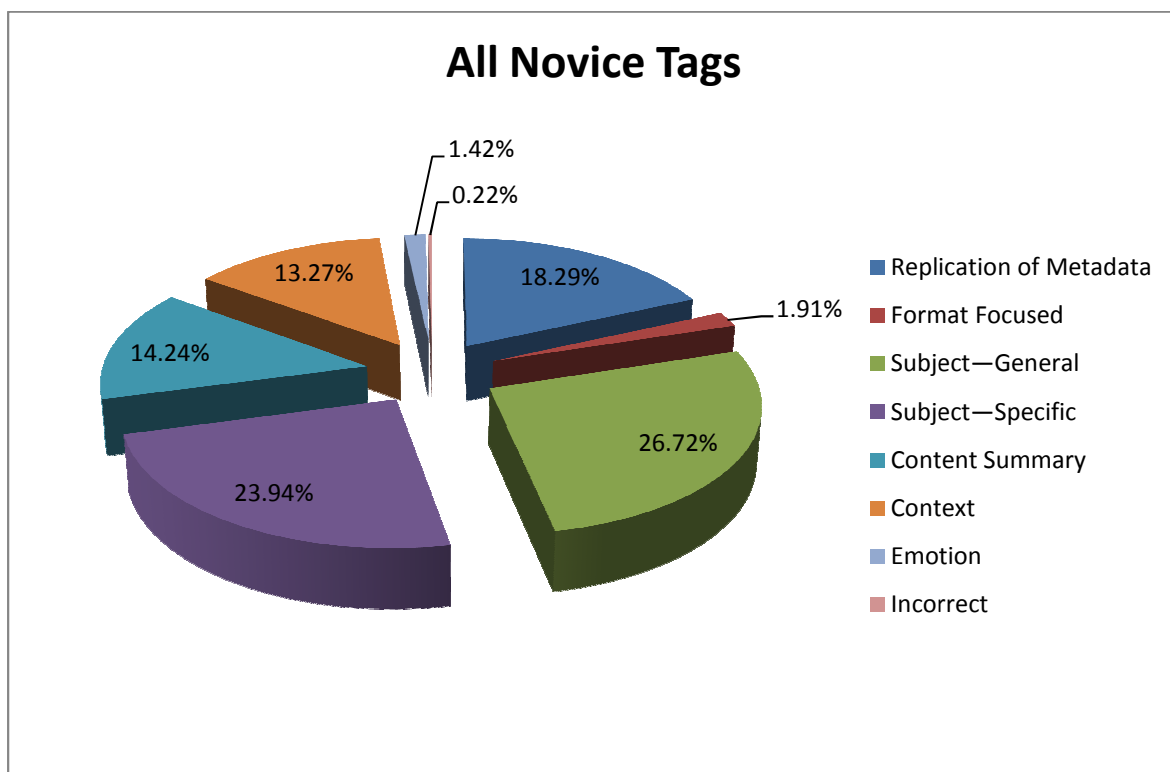
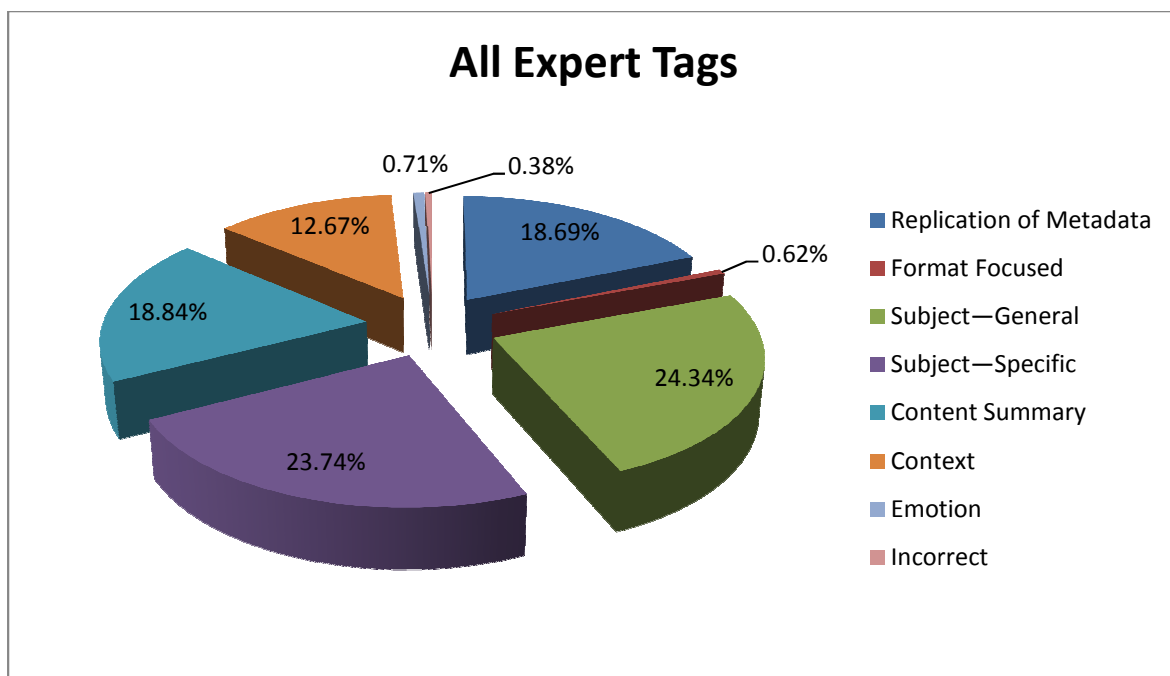


Figure 4.15 All Expert and Novice Tags by Category

A chi-square test for association was conducted between domain group (expert/novice) and tag category in order to test the significance between experts and novice tag difference for all items based on H₄.

H₄: The proportion of tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

All expected frequencies were greater than five. There was a statistically significant association between domain group and tag category, $\chi^2(7) = 77.149, p < .0005$.³⁹⁴ The association, however, is very weak, Cramer's V = 0.091.

Dividing the tags by format is necessary to best explore the similarities and differences between expert and novice tags. Photographs and documents illicit different responses from experts and novices (see Table 4.14 and Figure 4.16). Novices' photographic tags focused more on general subject terms while experts provided more content-summary and context tags for photographs through taking a broader view approach to the objects. Although experts accounted for more replication of metadata and incorrect tags than novices, the novices alone created format-focused photographic tags. These differences reflect the different approaches toward the photographs. Novices, having little domain knowledge background, attempt to identify individual parts of the photograph: a crowd, a library, a banner, a baton. Experts, on the other hand, identify what is going on in the captured scene: dissent, demonstration for racial justice, black-white solidarity.

³⁹⁴ The p-value is 1.5455×10^{-14} .

Experts created 396 unique photograph tags with novices creating 293 unique tags when compared to other tags within their domain groups. A cross-group comparison of unique tags finds an overlap of 116 tags, meaning 116 tags were created separately by both groups. The experts created 280 tags that the novices did not create, and the novices created 176 tags the experts did not create.

Table 4.14 Number and Percent of Expert and Novice Photograph Tags by Category

	Experts (n = 1705)		Novices (n = 2142)	
	No.	%	No.	%
Replication of Metadata	299	17.54%	277	12.93%
Format Focused	0	0.00%	89	4.15%
Subject—General	343	20.12%	628	29.32%
Subject—Specific	201	11.79%	266	12.42%
Content Summary	536	31.44%	515	24.04%
Context	292	17.13%	337	15.73%
Emotion	26	1.52%	28	1.31%
Incorrect	8	0.47%	2	0.09%

A chi-square test for association was conducted between domain group (expert/novice) and tag category in order to test the significance between experts and novice tag difference for photographs based on H_5 .

H_5 : The proportion of photographic tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

One cell in the chi-square test had an expected count of less than five; however, that cell's expected count was greater than one. Since it was the only expected count below five, the chi-squared analysis can still be run. There was a statistically significant

association between domain group and tag category, $\chi^2(7) = 142.043$, $p < .0005$.³⁹⁵ The association, however, is weak, Chramer's V = 0.192 (although stronger than the analysis of all tags).

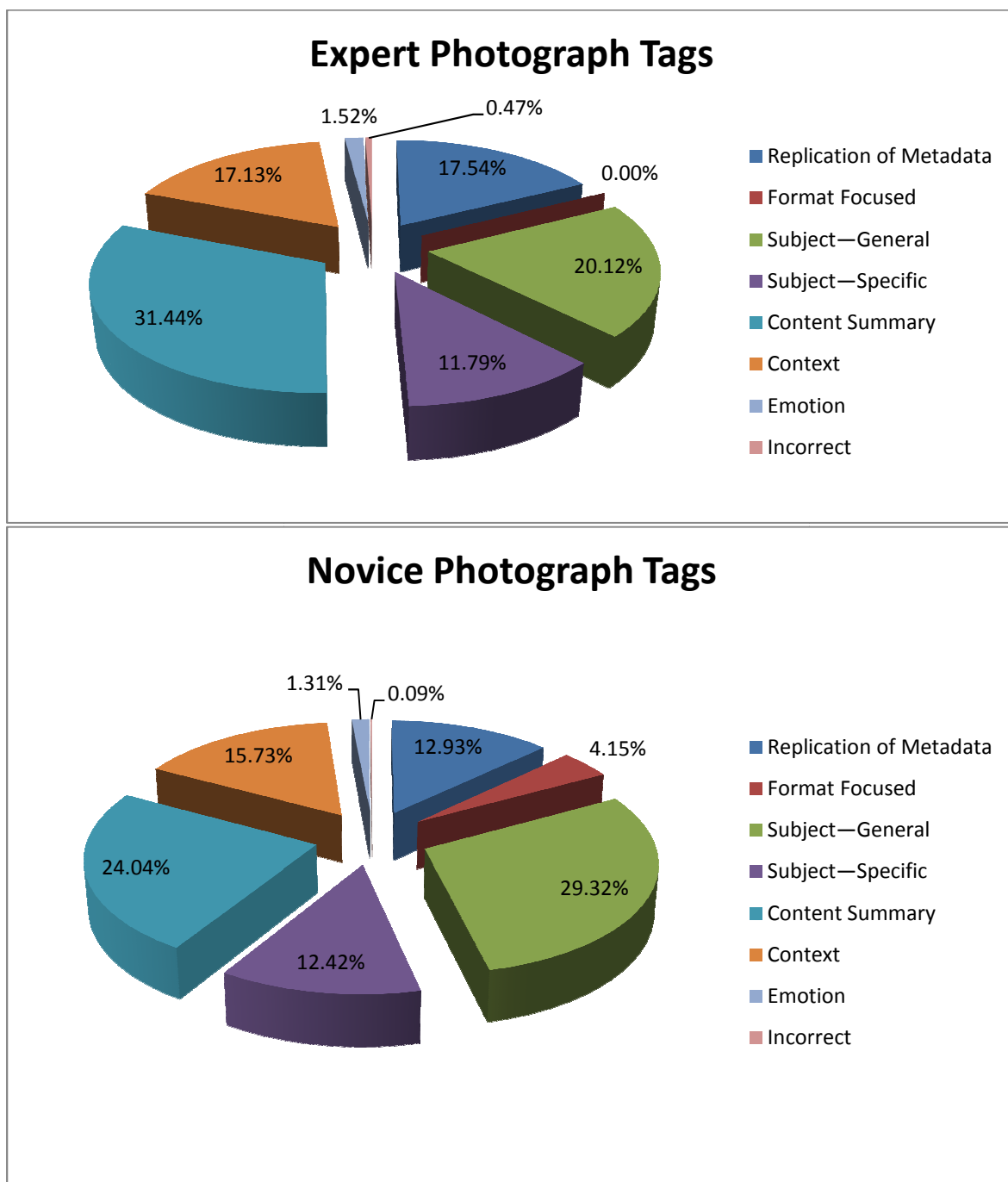


Figure 4.16 Expert and Novice Photograph Tags by Category

³⁹⁵ The p-value is 1.8965×10^{-27} .

The document tags offer a slightly different picture than the photographic tags (see Table 4.15 and Figure 4.17). In general, novices found the documents easier than photographs when it came to locating specific subjects since they only needed to extract from the text. This led to a 20-point increase in the subject—specific category for novices. At the same time, however, the novices reduced the number of content-summary tags by almost half and nearly eliminated format-focused tags in comparison with their photograph tags. A similar trend is seen with the expert tags as they increased subject—specific tags by 20 points while decreasing content-summary tags by 20 points. The experts did, however, include format-focused tags with the documents, unlike the photographs. Interestingly, the novices provided more context tags than experts for documents.

Table 4.15 Number and Percent of Expert and Novice Document Tags by Category

	Experts (n = 2494)		Novices (n = 2937)	
	No.	%	No.	%
Replication of Metadata	486	19.49%	652	22.20%
Format Focused	26	1.04%	8	0.27%
Subject—General	679	27.23%	729	24.82%
Subject—Specific	796	31.92%	950	32.35%
Content Summary	255	10.22%	208	7.08%
Context	240	9.62%	337	11.47%
Emotion	4	0.16%	44	1.50%
Incorrect	8	0.32%	9	0.31%

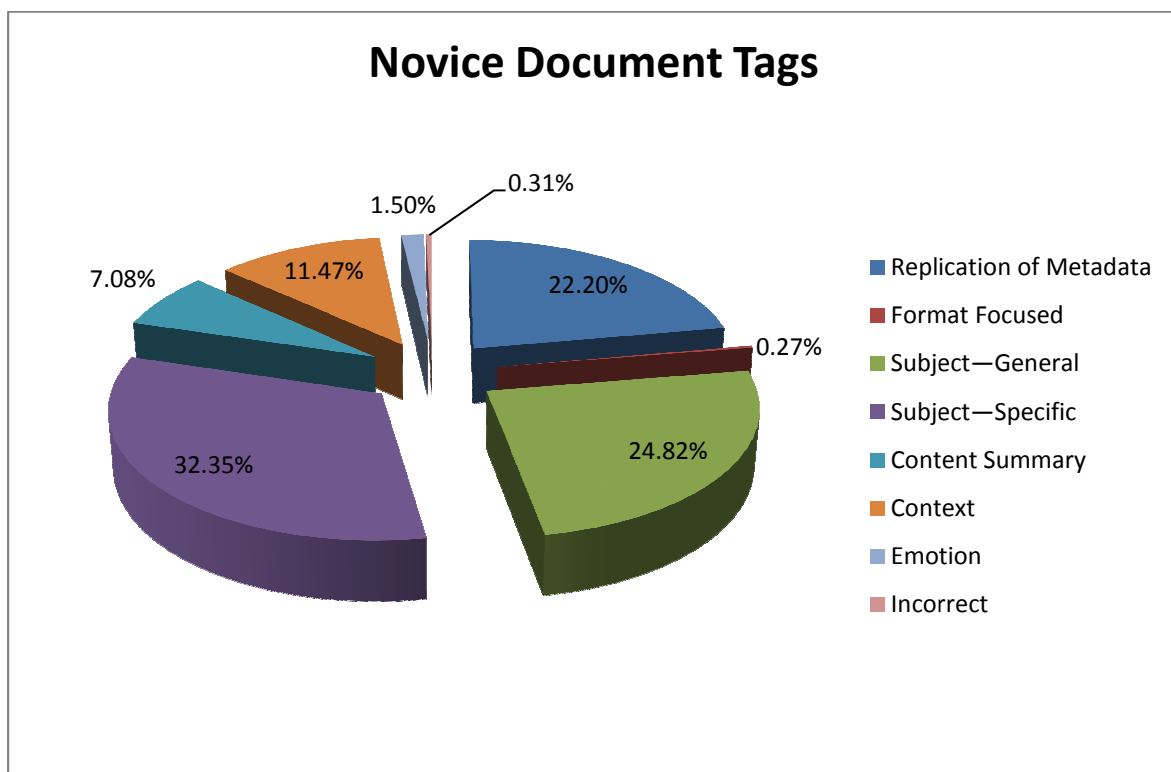
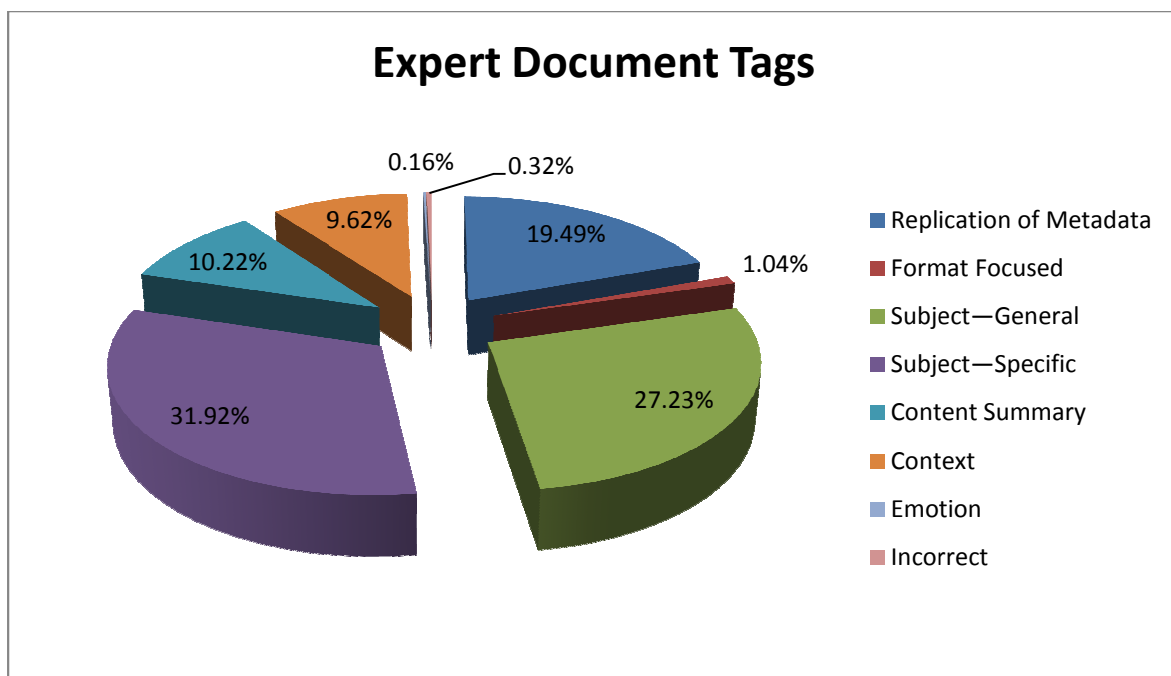


Figure 4.17 Expert and Novice Document Tags by Category

When compared within their own domain groupings, the experts created more unique tags (685) than the novices (579). A cross-group comparison of unique tags found 295 terms in both groups' unique tag lists. The experts created 404 unique tags which the novices did not create, while the novices created 294 unique tags that the experts did not produce.

A chi-square test for association was conducted between domain group (expert/novice) and tag category in order to test the significance between experts and novice tag difference for documents based on H_6 .

H_6 : The proportion of document tags in each coding category in a minimally processed digital archive is affected by a user's domain knowledge.

All expected frequencies were greater than five. There was a statistically significant association between domain group and tag category, $\chi^2(7) = 67.889, p < .0005$.³⁹⁶ The association, however, is weak, Chromer's $V = 0.112$ (although stronger than the analysis of all tags, but weaker than the photograph tags).

All three tested hypotheses for RQ 1(a) indicated a statistically significant association between domain group and coded tag category. The associations are all relatively weak based on low Chromer's V values of 0.091 (H_4), 0.192 (H_5), and 0.112 (H_6). The small differences between domain groups likely caused the low level of associative strength. The proportion of tags within several categories, such as replication of metadata, was consistently close between both experts and novices thereby limiting the strength of statistical association. Increasing the number of participants (and therefore

³⁹⁶ The p-value is 3.941×10^{-12} .

increasing the number of tags) could see the categorical differentials increase and strengthen the statistical association.

4.2 Research Question 1(b): Are there differences between expert and novice users' opinions of the tagging experience and tag creation considerations?

All participants completed a post-questionnaire with close-ended and open-ended questions. The questions were designed to identify differences between expert and novice users' opinion of the tagging experience and what the participants considered during tag creation. Participants indicated their structured questions' responses on a 5-point Likert scale with additional information for each category in open-ended questions. Mann-Whitney U tests were run for each set of questions to determine if the differences in responses between experts and novices (if any) were statistically significant. Mann-Whitney U tests were selected since the dependent variable in each case was an ordinal variable (5-point Likert scale), and t-tests require a continuous dependent variable.

Participants' indication of agreement with three statements (1 = Strongly Disagree; 5 = Strongly Agree) explored the participants' opinions of the tagging experience. Table 4.16 summarizes the findings of the Mann-Whitney U tests with the first statement's scores being similarly distributed (thereby reporting the comparison of medians) and the final two statement scores being not similarly distributed (thereby reporting the mean rank) based on H_7 - H_9 .

H_7 - H_9 : Expert and novice users' opinions of the tagging experience are different for ease of tagging in general (H_7); difficulty in tagging documents compared to photographs (H_8); and difficulty in tagging photographs compared to documents (H_9).

The tests found no statistical significant differences between experts and novices for all three statements, meaning the participants shared similar experiences tagging. Additionally, this negates the tagging experience from affecting tag creation in the study. Exploring the composite mean scores of each statement highlights further similarities. All participants scored the positive statement relatively high (4.22). Both experts and novices did not think one format was more difficult than the other, reporting a combined average of 2.87 for documents being more difficult than photographs, and 2.62 for the reverse.

Table 4.16 Mann-Whitney U Test Comparison of Tagging Experience Statements

Statement	Grp	Mean	Median	U	z	p
I found submitting tags easy	Exp	4.20	4.00	471.0	.346	.729
	Nov	4.23	4.00			
Score distribution for statements below found not similar based on visual inspection			Mean Rank			
I found tagging documents more difficult than tagging photographs	Exp	2.73	28.77	502.0	.794	.427
	Nov	3.00	32.23			
I found tagging photographs more difficult than tagging documents	Exp	2.47	28.67	505.0	.841	.401
	Nov	2.77	32.33			

Responses to the open-ended question, “What would make the tagging experience better?” also showed similarities in the experiences of experts and novices. The majority of the combined (expert and novice) responses (41.7%) indicated some frustration with the tagging system itself. Participants wanted better methods for adding or manipulating tags, such as sorting features (E03), spellcheck option (N08, N06, N14), easier methods for reusing others’ tags or voting for someone else’s tag (E13, E17, E21, N11, N18, N22, N24, N27, N29), and providing a premade set of approved tags for use in the collection (N13, N17). While most participants found the ability to see others’ tags useful, one participant preferred not seeing others’ tags since, “It influences my tags” (N26). Finally,

many participants wanted a more intuitive tagging system that was both easier to locate on the page, and allowed for more viewing options for the item being tagged (E08, E10, E11, E19, E30, N01, N05, N06, N25).

A small number of all participants (11.7%) had issues with the scanned versions of the documents (particularly the handwritten ones), and one participant (E07) wanted to know more about the end users to understand the purpose of the tags better. An equal number of experts and novices (5 of each or 16.7% of responses) thought more background information would make the tagging experience better. Three participants wanted the system to provide more metadata for the items (E12, E26, N19). The other experts' comments were vaguer on the type of information desired (E01, E18, E24), while the novices specifically mentioned the need for more personal background on the collection's subject (N02, N03, N15, N19, N30). Participant N02 stated, "I enjoyed the tagging experience, but it would have been easier if I had known more about the civil rights movement in Milwaukee, especially where the pictures are concerned." Similarly, participant N30 stated, "I have very little background knowledge regarding the collection/events that transpired, so I felt that many of my tags lacked the depth needed to differentiate the items from one another."

Tag creation considerations were investigated through expert and novice users' agreement/disagreement with five statements (1 = Strongly Disagree, 5 = Strongly Agree). Similar to the above, the analysis used Mann-Whitney U tests rather than t-tests since the data were ordinal rather than continuous. Table 4.17 summarizes the findings of the Mann-Whitney U tests with the first three statements' scores being not similarly

distributed (thereby reporting the comparison of medians) and the final two statement scores being similarly distributed (thereby reporting the mean rank) based on H_8 - H_{14} .

H_8 - H_{14} : Expert and novice users' opinions of the considerations for the creation of tags are different for how others would find the item (H_{10}); how the tagger (user) would find the item (H_{11}); the content of the tagged item (H_{12}); the format of the tagged item (H_{13}); and other users' tags (H_{14}).

Novice participants' agreement (mean rank = 34.83) with the statement, "How others would find the item," at a statistically significantly higher level than expert participants' (mean rank = 34.83), $U = 580$, $z = 2.132$, $p = .033$.

Table 4.17 Mann-Whitney U Test Comparison of Tagging Consideration Statements

Statement	Grp.	Mean	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
How others would find the item	Exp	4.10	26.17	580.0	2.132	.033*
	Nov	4.60	34.83			
The content of the item	Exp	4.50	29.50	480.0	.531	.596
	Nov	4.67	31.50			
The format of the tagged item	Exp	3.33	31.03	434.0	-.245	.806
	Nov	3.23	29.97			
<i>Score distribution for statements below found similar based on visual inspection</i>			Median			
How I would find the item	Exp	4.27	4.00	554.0	1.728	.084
	Nov	4.60	5.00			
Other user's tags	Exp	3.63	4.00	529.0	1.279	.201
	Nov	3.90	4.00			
* Indicates statistically significant findings						

The tests found no statistically significant difference between expert and novice participants for the remaining statements, meaning the participants considered each of the elements similarly. Viewed in the aggregate, the participant means indicate only two other elements with high agreement (above 4.0) in addition to the statistically significant

finding: How I would find the item (4.43) and the content of the item (4.58). The format of the tagged item averaged 3.28 indicating a slightly higher than average agreement.

The open-ended questions following this set of statements ask participants for any additional considerations they used while creating tags and if their considerations changed during the tagging process. Similar to the tagging experience question, both expert and novice participants shared similar opinions to the creation of tag considerations. Four experts and six novices (16.7% of all participants) stated some alternations in tagging considerations as they progressed through the collection. The emotional connection to the materials caused the change for three experts, one of whom stated, “The tags became more emotionally connotative as I progressed through the sequence of items” (E05). A second expert (E27) said s/he returned to previous items to add more tags once the collection began reawakening memories from the past. Additionally, one of the novices described the change from general tags to more detailed tags (N19).

Seven experts and seven novices (23.3% of all participants) indicated some broader concern over keywords or subject content of their tags. Two experts (E10 and E11) tried to be as descriptive as possible, while a novice (N24) purposely created broader subject-based tags. Yet another participant (E02) actively created both general and specific tags. Novice 12 looked for “unusual words” within the documents and another novice, “tried to use tags that differentiated the items from one another” (N30).

An additional consideration for some participants was the potentially controversial language used within some of the documents. One expert stated, “I thought

about political correctness, the use of certain tags (is ‘white people’ or ‘black people’ helpful?), trying to keep my personal response out of the tag (e.g., don’t include ‘your’ or ‘grammatically incorrect’ or ‘ignorant’ as tags because they are [sic] my biased opinions and are not helpful for searching” (E17). Another participant considered “whether it was proper to use a tag that no one uses (an ethnophaulism) if it is a direct quote or whether this might be offensive to use as a tag even if a quoted word” (E07). Finally, several participants took accuracy and consistency under consideration while creating their tags (3 experts, 2 novices, 8.3% of all participants).

4.3 Research Question 2(a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with metadata in a traditionally processed digital archive?

One of the goals of including user-generated tags as supplemental metadata within a minimally processed digital archive is the potential for replicating or replacing the detailed item-level metadata found in traditionally processed digital archives. The dissertation explores this possibility through using a test collection sampled from an existing collection, thereby allowing both the presentation of minimal metadata for the experiment and extracting the full item-level metadata for comparison with the user-generated tags. The full item-level metadata not included in the minimally processed metadata seen by participants, called unselected metadata, was aggregated into two lists (photographs and documents) for comparison with the participant created tags. Although research question 2(b) tests for an association between prior domain knowledge and the proportion of tags that match the unselected metadata below, it is first important to highlight the ways in which tags generated by both experts and novices in a minimally

processed collection correspond with the metadata of a traditional item-level processed digital archive.

The Dublin Core metadata standard remains a primary choice for digital collections due to its flexible interoperable nature. As such, it can also serve as a categorical structure for highlighting the similarities and differences between tags corresponding with existing metadata. The *March on Milwaukee* uses different combinations of the majority of the fifteen Dublin Core elements within its metadata template. Within the Groppi Papers, the existing collection uses the following elements: Title, Creator, Subject, Description, Publisher, Date, Type, Format, Identifier, Language, Relation, and Rights. Table 4.18 displays the different unique field names mapped to Dublin Core elements for both documents and photographs within the existing collection. Several of the fields were included within the minimal metadata provided to participants, as indicated with an asterisk (*) in the table. Although the title field was included in the minimal metadata, the titles used in the experiment were generalized (e.g., Photograph 1, Support Mail 1, etc.), whereas the existing collection's titles were item-level specific (e.g., James Groppi and Vel Phillips on school bus, circa 1967-1968).

Table 4.18 Existing Metadata Template for Groppi Papers

Dublin Core Element	Unique Field Names	
	Photographs	Documents
Title	Title*	Title*
Creator	Creator*	Creator*
	Photographer	
Subject	Subject	Subject
	Topic	Topic
	Keywords	Keywords
	People	People
	Organization	Organization
	Event	Event

	Place	
Description	Description	
Publisher	Digital Publisher*	Digital Publisher*
Date	Date	Date
	Date Digitized*	Date Digitized*
Type	Type (DCMI)*	Type (DCMI)*
Format	Original Item Type*	Original Item Type*
	Original Item Format	Original Item Format
	Digital Format*	Genre
		Digital Format*
Identifier	Original Item ID	
	Digital ID	
	WHS Image ID	
Language		Language
Relation	Original Collection*	Original Collection*
	Repository*	Original Item Location*
	Digital Collection*	Folder Title
	Part of*	Repository*
		Digital Collection*
Rights	Rights*	Rights*

(* indicates field included in minimal metadata presented to participants)

Aggregated lists of the so-called unselected metadata, that is the item-level metadata from the existing collection not included in the sample collection used in the experiment, were compiled for six Dublin Core elements: Title, Date, Description, Subject, Identifier, and Format. The lists were first made based on format (photograph, document) and then merged into a combined list for comparison with the user-generated tags. Table 4.19 lists the number of metadata terms within each format and element grouping. The documents did not contain any description or identifier metadata.

Table 4.19 Number of Unselected Metadata Terms by Dublin Core Element

	Title	Date	Description	Subject	Identifier	Format
Photographs	61	7	165	68	38	2
Documents	37	12	0	50	0	5

The unselected metadata terms were compared to the expert and novice tags initially by format and subsequently as complete sets. Table 4.20 reports the number and percent of matching terms for each format and element grouping. As a whole, the numbers suggesting a high level of tags matched the unselected metadata for the title and subject elements, while metadata from the date and format fields did not usually match. Additionally, the identifier metadata never matched across the entire sample collection's tags, suggesting it would be a poor metadata field to expect user-generated content to match. This is not surprising since the identifier is typically only known to the repository itself, and not generally seen on the digital object. The description field, which only occurs for the photographs, was nearly twice more likely matched with an expert's tag than with a novice's.

Table 4.20 Number and Percent of Unselected Metadata Terms Matching User-Generated Tags by Dublin Core Element

	Photographs				Documents			
	Expert		Novice		Expert		Novice	
	#	%	#	%	#	%	#	%
Title	52	85.2%	34	55.7%	28	75.7%	20	54.1%
Date	1	14.3%	1	14.3%	3	25%	3	25%
Description	68	41.2%	44	26.7%	n/a		n/a	
Subject	43	63.2%	36	53%	35	70%	19	38%
Identifier	0	0%	0	0%	n/a		n/a	
Format	0	0%	0	0%	2	40%	2	40%

Although the number of tags matching unselected metadata does illuminate some similarities and differences between expert and novice tags, further comparison requires focusing on the tags themselves. The following section discusses the matching tags for each element set unique to each domain group by format grouping. Table 4.21

summarizes the percent of unique matching tags for each domain, format, and element grouping.

Table 4.21 Percent of Tags Matching Unselected Metadata Unique by Dublin Core Element

	Photographs		Documents	
	Expert	Novice	Expert	Novice
	%	%	%	%
Title	36.5%	2.9%	28.6%	0%
Date	100%	100%	0%	0%
Description	39.7%	6.8%	n/a	n/a
Subject	25.6%	11.1%	14.3%	21.1%
Identifier	0%	0%	n/a	n/a
Format	0%	0%	0%	0%

The photographs best highlight the difference between the expert and novice unselected metadata matching tags. In four elements (title, date, description & subject), both experts and novices provided at least one tag that matched the unselected metadata but was not included in their counterpart's tags. Although both domain groups (expert, novice) created these unique tags, the experts did so at a much higher rate. Within the title element metadata, for example, experts had fifty-two total tags match unselected metadata with thirty-four for the novice tags. Of these tags, thirty-three were duplicated by both experts and novices. The experts tag set included nineteen matching tags that were not in the novice set, while the novices only created a single additional unique tag. Focusing on the tags themselves, the unique expert tags provided specific information or identification of things within the images, such as St. Boniface, Vel Phillips, and Madison. It is also interesting to note the unselected metadata that was not replicated by any tags included general words, such as "back" or "between" which are difficult to include within tags unless using a compound, multiword, or phrase tag. The title non-

replicated unselected metadata also included date tags (1965, 1966, 1968) which were difficult for participants to identify within a photograph, given no additional clues. This trend is duplicated with the date-element-specific metadata, and the low matching rate. In fact, the two matching tags within the date element are the same two dates (1969 and 1967) which were unique matching tags within the title element for both expert and novices.

The final two elements with tags matching unselected metadata within the photographs, description and subject, offer similar similarities and differences as stated above. Within the description element, forty-one matching tags were shared by both domain groups, with the experts providing twenty-seven additional matching tags and the novices just three. These unique tags included both specific terms, such as 1967 (novice) and Wisconsin (expert) as well as general terms, such as small (expert) and people (novice). The description element unselected metadata included 188 terms that did not match any tags. Although many of these metadata were again more general in nature, several provided specific information not recognized by the participants, including Bishop Athieliski, Harold Froehlich, and Howard Berliant.³⁹⁷ Within the subject element, thirty-two tags that matched unselected metadata were shared by both domain groups, with novices creating an additional four and experts an additional eleven tags. The unique tags echo the previous discussion with specific and general terms. For the subject element, 21 metadata terms were not matched by participant tags; however, most were rather innocuous and one could reasonably assume they might be replicated given enough tag development over time (e.g., activists, arrests, courts, law, etc.).

³⁹⁷ Harold Froehlich was a state representative standing next to Fr. Groppi in a photograph, and Howard Berliant was a photographer for one of the images in the sample collection.

The trends noted within the photographs do not continue with the document tags. Unlike the photographs, the documents only had unique tags matching unselected metadata within the title and subject elements (all generated by experts). Furthermore, the unique document tags do not provide meaningful additional information. In the title element, for example, experts created eight unique tags (1, 3, 5, 20, 26, 31, 6, and June). Although these look like simple numbers, they are parts of dates used within the titles for the letters. The experts tended to provide the full date (June 4, 1969) whereas novices usually provided an abbreviated date (1969). Within the subject element, the five additional expert tags matching metadata were active terms (e.g., non-violence, struggle, etc.) whereas the four unique novice tags were more passive descriptive terms (e.g., whiteness, relations, etc.). Although these minor differences exist, the participants primarily shared matching tag terms for documents across all elements with forty title, three date, sixty description, and two format tags being shared.

The unselected metadata not replicated with the documents continues the trend of the photographs, with limited amounts of key information included within the non-replicated terms. The format element metadata for both photographs and documents did not match well with participants' tags, with only two of a possible seven terms matching. The lack of replication, in this case, is primarily due to the archival language used to describe formats. The seven unselected metadata terms (photographic, prints, letters, manuscripts, typescripts, handwriting, correspondence) were, in fact, all included within the participants' tags but with different expressions. While none of the participants used typescripts, they did include typewritten; likewise for handwriting, where participants did include handwritten.

4.4 Research Question 2(b): Does user knowledge affect the proportion of tags matching unselected metadata in a minimally processed digital archive?

One of the goals of including user-generated tags as supplemental metadata within a minimally processed digital archive is the potential for replicating/replacing the detailed metadata that is not included within minimal processing. The dissertation explores this possibility through using a sample collection from an existing collection, thereby allowing a comparison of the users' tags and the unselected metadata. A compiled list of the full metadata for the sample items by format was compared to the minimally processed metadata provided to users. The results created two lists of unselected metadata with the photograph list containing 278 terms and the document list containing 150 terms. The unselected metadata was compared to the lists of unique tags by domain and format, generating a table of matching and non-matching counts (see Table 4.22); figure 4.18 illustrates these differences.

Table 4.22 Proportion of Tags Matching Unselected Metadata

		# Match	% Match	# Non-match	% Non-match
Photographs	Expert	95	34.17%	183	65.83%
	Novice	70	25.18%	208	74.82%
	Combined	102	36.69%	176	63.31%
Documents	Expert	80	53.33%	70	46.67%
	Novice	70	46.67%	80	53.33%
	Combined	86	57.33%	64	42.67%

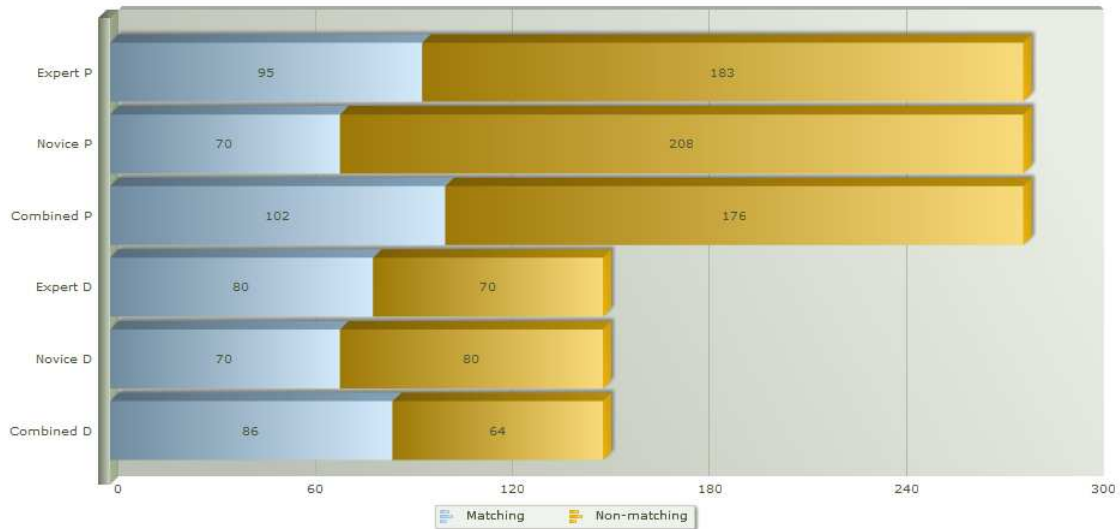


Figure 4.18 Proportions of Matching/Non-Matching of Tags to Unselected Metadata

For both the photographs and documents, the experts' tags replicated the unselected metadata more than novices'. Not surprisingly, however, the highest matching rate for both formats occurred with the combination of experts' and novices' tags. A chi-square analysis of the data was conducted to test if there was a statistically significant association between the number of matching tags and the user's domain knowledge based on H_{15} .

H_{15} : The proportion of tags matching unselected metadata is affected by the user's domain knowledge.

Individual chi-square tests were run for the photograph and document data. In both tests, all expected cell frequencies were greater than five. The photograph test found a statistically significant association between the user's domain knowledge group (expert or novice) and the proportion of tags matching existing metadata, $\chi^2(1) = 5.386, p = .020$.

The association, however, is weak at best, $\phi = 0.098$, $p = .020$. The document test, however, did not find a statistically significant association between the user's domain knowledge group and the proportion of tags matching existing metadata, $\chi^2(1) = 1.333$, $p = .248$. Therefore, the hypothesis is rejected in the case of documents, but accepted for photographs, with the preface that the association is very weak. The weak association indicates the difference between experts and novices remains quite close. Similar to previous weak associations, increasing the sample size might increase the associative strength.

4.5 Research Question 3(a): In what ways do tags generated by expert and/or novice users in a minimally processed collection correspond with existing users' search terms in a digital archive?

Social tags cannot serve as useful tools if they do not assist with other users' information retrieval. Similar to the previous research question, the use of a sample from an existing collection provides the necessary data for comparing tags with existing query terms. The Digital Collections at the UWM Libraries provided the query logs for the month of January 2014. Parsing of the server logs resulted in 59,325 unique query terms used to search across all collections hosted by UWM-DC. Further reduction by collection-specific searches found 1,609 unique query terms used to search the *March on Milwaukee* collection alone. Tables 4.23 and 4.24 display the results of comparisons for both query lists to the unique tag terms created by experts, novices, and both groups combined. Table 4.24 also includes a comparison with the unselected metadata for both photographs and documents compiled for the previous research question.

Table 4.23 Comparison of All Collection Query Terms and Tags

	# Match	% Match	# Non-match	% Non-match
Expert	575	0.97%	58,750	99.03%
Novice	442	0.75%	58,883	99.25%
Combined	694	1.17%	58,631	98.83%

Table 4.24 Comparison of *March on Milwaukee* Query Terms and Tags

	# Match	% Match	# Non-match	% Non-match
Expert	333	20.70%	1,276	79.30%
Novice	243	15.10%	1,366	84.90%
Combined	360	22.37%	1,249	76.63%
Unselected Metadata	398	24.74%	1,211	75.26%

An examination of all of the matching tags/metadata terms highlights the relationship between expert tags, novice tags, and metadata terms. Figure 4.19 illustrates the relationships in a Venn diagram with the number of unique matching terms indicated for each segment and examples of terms found in each segment. The metadata segment is used for the unselected metadata grouping; for example, the Venn diagram segment overlapping expert and metadata show 49 unique terms that matched the query term list occurred within both the expert and unselected metadata lists.

As noted in the middle of the diagram, 129 terms were included in all three groups (expert, novice, and metadata). The diagram did not provide enough room for examples of this particular subgrouping. Many of the terms included in all three groups describe major themes of the collection as well as key persons or places from the collection. Examples of theme-related terms include: black, bus or bussing, colored, demonstration(s), housing, march or marching, protest, power, integration, segregation,

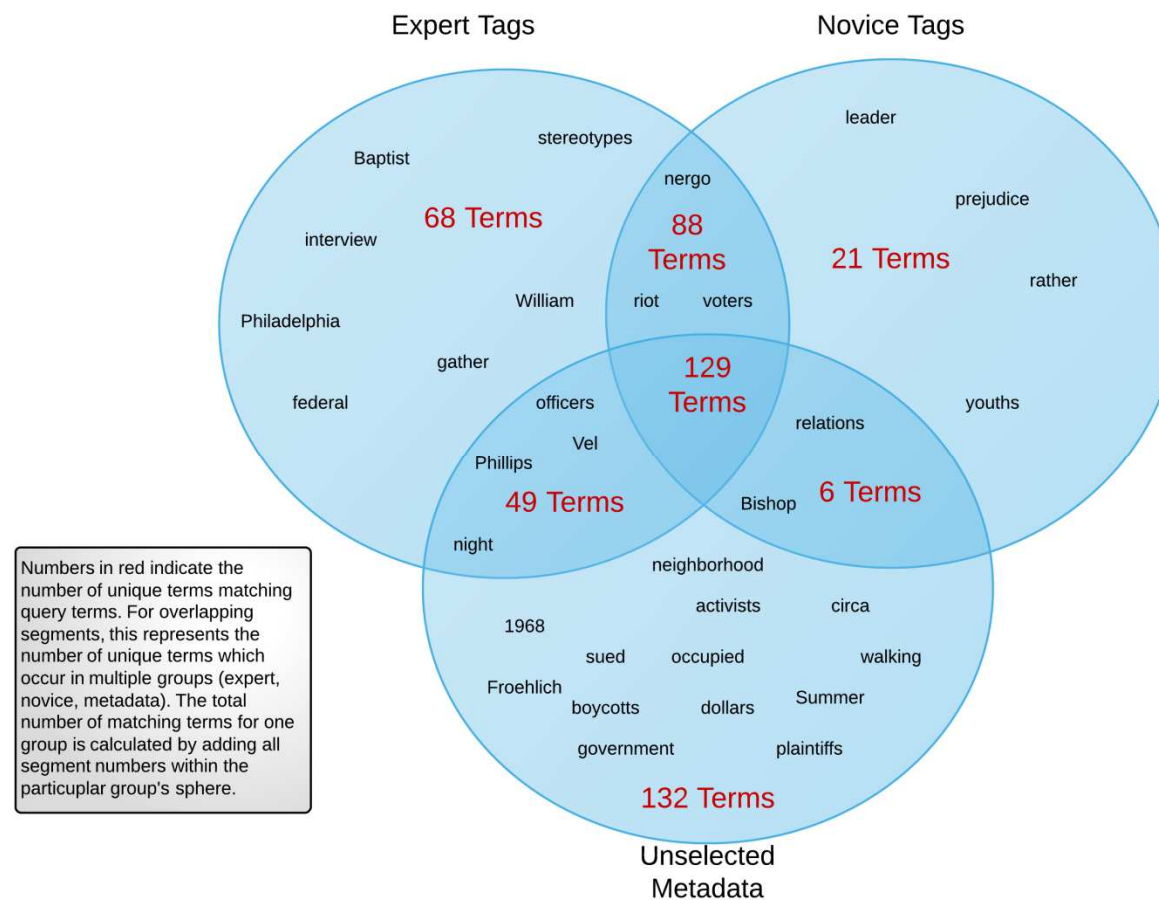


Figure 4.19 Tags and Unselected Metadata Matching User Query Terms Venn Diagram

school(s), and youth. Other terms highlight important elements or icons of the photographs, such as “burning” for the image of the Freedom House burning, “fist” for the image of Groppi’s raised fist of resistance, and “wagon” for the image of an arrested Fr. Groppi sitting in a police wagon. Several dates, or parts of dates, appeared in the shared list as well, including 1966, 1967, December, February, March, May, July, August, and September. A final characteristic of this subgrouping of terms is the inclusion of key people or places from the photographs and documents. Examples include groups like the Commandos and the NAACP, important places, such as Milwaukee and Wisconsin, and authors or subjects of the letters and photographs, such as Groppi himself, LaValle, Crooms, McKissick, Waiss, and Waverly. The inclusion of all of the subgroupings’ terms by experts, novices and the unselected metadata indicate their importance to both the collection and users’ perception of the collection.

An analysis of the participant-exclusive tags matching user query terms also notes some important themes and potential causality (looking at expert only, novice only, and expert and novice subgroupings combined). Many of the tags are different forms, versions or conjugations of words found within the metadata terms. Often it is simply a plural version, such as newspaper appearing in the metadata, expert, and novice subgrouping while newspapers is only in the expert subgrouping (additional examples will include associated subgroupings in parenthesis). Additional examples are youth (metadata, expert & novice) and youths (novice only), and group (metadata, expert & novice) with groups (expert only). More often, however, the tag is a different version, such as desegregation (expert & novice) versus de-segregation (metadata, expert & novice). In addition, taking the alterations yet further, some of the participants’ tags

conjugate the term to desegregate (novice only), creating another variation. Finally, the tags offer abbreviations for terms or phrases, such as “Rev” for Reverend, “feb” for February, or “photos” for photographs.

Although the differences between these tags and the metadata terms appear minor, the matching between user search terms and the alternative variations raises their importance and significance. Modern users have become accustomed to the Google style search that automatically corrects misspellings and searches multiple tenses, cases, and even derivations of the words, whereas most content management systems for digital collections, such as CONTENTdm, do not make such adjustments to search terms. The inclusion of the term variations within the query log indicates users are still searching with vernacular, and the participants’ tags also containing similar variations allow for successful matching between tag and query terms.

Additional analysis of the participants’ matching tags not included within the metadata reveals another trend, the importance and/or usefulness of transcription of documents. The vast majority of these tags come from the document tags rather than the photographic tags. Specifically, 102 tags occurred only within the document tag sets and an additional thirty-six tags occurred within both the photograph and document sets. This represents a combined 78% of the 177 tags which match user query terms but do not match unselected metadata (or 57.6% if excluding the tags also occurring within the photograph sets). When looked at by domain knowledge group, the unique tags created by experts alone or novices alone are consistent with 67.6% and 66.7% respectively (unique tags occurring in both expert and novice groups raises the percentage to 88.6%). Since the document unselected metadata does not include the description Dublin Core

element, it also does not contain transcribed information from the documents themselves. The tags, on the other hand, often did come from the document contents, and the above analysis suggests a strong connection between the tags and user search terms.

4.6 Research Question 3(b): Does user knowledge affect the proportion of tags matching query terms in a minimally processed digital archive?

Expert users' tags match the two query term lists in higher proportions than the novices'; however, the combination of tags outperformed both individual groupings. Chi-square analysis of the data was performed to test for a statistically significant association between users' domain knowledge grouping (expert, novice) and the proportion of tags terms that matched both query-log term lists based on H_{11} :

H₁₆: The proportion of tag terms matching users' query log terms is affected by user's domain knowledge.

Individual chi-square tests were run for the all-collections query list and the *March on Milwaukee*-specific query list. In both tests, all expected cell frequencies were greater than five. The all-collections test found a statistically significant association between the user's domain knowledge group and the proportion of tags matching query terms, $\chi^2(1) = 17.826, p < .0005$.³⁹⁸ The association, however, is weak at best, $\phi = -0.012, p < .0005$. The *March on Milwaukee*-specific test found a statistically significant association between the user's domain knowledge group and the proportion of tags matching query terms, $\chi^2(1) = 17.128, p < .0005$.³⁹⁹ The association, however, is weak at best, $\phi = 0.073, p < .0005$. Both weak association findings replicate issues noted with

³⁹⁸ The p-value is 0.000024.

³⁹⁹ The p-value is 0.000035.

earlier statistical tests. Although there are statistical differences between experts and novices, the differences are minor with the groups performing close to each other.

Increasing the sample size could increase the difference between experts and novices, thereby strengthening the statistical associations.

CHAPTER FIVE: DISCUSSION

The findings for each of the three main research questions highlighted minute differences between expert and novice participants' tags. Although differences exist, in all cases the differences were either statistically insignificant or a very weak association with the domain knowledge group. The data shed light on several areas of both practical and theoretical implications. This chapter discusses the theoretical implications, practical implications, methodological implications, and limitations of the dissertation's results.

5.1 Theoretical Implications

The archival backlog problem of the past twenty years arose from the emergence of the postmodern movement's increase in the number and types of collections within repositories. The rise in popularity and practice of minimal processing served as a direct response, and created the right conditions for social tagging's role within digital archives. Although social tagging offers several practical benefits (discussed later in the chapter), the combination of minimal process and social tags have significant theoretical implications, specifically adherence to the postmodern ideals.

Archival postmodernism reacted to the limited voices and perspectives represented in archival collections during the post-World War II collecting spree. Following initial calls for archivists' active role in identifying and filling collection gaps in the 1970s, several archivists suggested the need for including outside voices within archival description (in addition to the added collecting emphasis). Chapter Two further outlined these developments, and social tagging's potential role as supplemental metadata and archival description. The dissertation's results reinforce these possibilities.

The dissertation study fits the postmodern requirements for heterogeneous description through the inclusion of a wide range of participants. Both the expert and novice groups included multiple generations, ethnic groups, religious or non-religious affiliations, regional locations, and educational levels. Unlike other participatory archives and archival tagging studies, the dissertation limited the influence of the archival voice through reliance on minimal processing. This allowed unfettered development of differing opinions, interpretations, and descriptions of the thirty records.

Focusing on the content summary, context, and emotional tags, the participants successfully produced tag-based expressions of their unique perspectives. Each participant based their tags on their own understanding of the materials. Through looking at the aggregate of both domain groups of tags, one can see a conversation develop as some tags increased in use and popularity. However, even tags without replication by others provided additional information for potential researchers and archival users. Findings of the limited sample of the dissertation suggest the combination of social tags and minimal processing in digital archives would successfully produce a postmodern digital archives.

The inclusion of user-generated description can also be seen as a step toward a more democratized archive through titling the archivist/user power dynamic further toward the user. Through actively engaging in the archival process, social tagging allows users to further claim ownership and agency over the records. Tagging provides an avenue for users' identification of value during the selection of items to tag and the words used during the tagging process. The power shift toward users combined with the

heterogeneous description places social tagging comfortably within a postmodern archival worldview.

In addition to the postmodern implications of the study, the dissertation results reinforce or broaden the findings of previous archival and social tagging studies, specifically focused on tagging behavior and the nature of social tags. Previous participatory archival research often focused on descriptions of the potential benefits of user participation or engagement rather than empirical testing. Studies by Flinn, Eveleigh, or Huvila, for example, encourage the expansion of archival engagement through public collaboration throughout the archival processes.⁴⁰⁰ Although these previous studies occasionally use case studies in their arguments or discussion, the lack of empirical evidence supporting the benefits of participatory models for archives caused some pushback from both the archival community and others. The dissertation study's findings offer the needed evidence demonstrating the benefits of allowing users with a broad range of backgrounds into the description processes through providing social tags. The resulting tags add the multiple diverse interpretations of the archival materials suggested by participatory archival research. Furthermore, the findings also reinforce Evans' discussion of relieving archives of the temporal and fiscal burdens of increased collections through "acting as partners" or "organizing agents" with users for the item-level description.⁴⁰¹

⁴⁰⁰ Flinn, "An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge."; Eveleigh, "Welcoming the World: An Exploration of Participatory Archives"; and Huvila, "Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management,"

⁴⁰¹ Evans, "Archives of the People, by the People, for the People," 397.

The dissertation findings also answer calls for additional research into the content created by users, and specifically how it could be integrated or supplement archival description.⁴⁰² The successful matching of participants' tags with both the unselected metadata and the query terms suggests social tags are an effective additional or supplemental access point to the digital archives. The lack of incorrect tags within the study's findings also reinforces Palmer's argument to treat users as "peer collaborators...rather than outside interlopers."⁴⁰³

The dissertation also provides theoretical implications based on previous research into social tagging in general and social tagging within archives specifically. The comparison of participants' tags with the unselected metadata, and the high degree of successful matches replicate the previous findings of Kipp and Campbell, who found tags often develop the same concepts as traditional indexing, although in this case through metadata rather than index terms.⁴⁰⁴ Participants' wish for more direct appeals and guidance on desired tags from archives combined with the results' limited instances of incorrect tags echoes Guy and Tonkin's previous suggestions on improving tagging behavior and conditions.⁴⁰⁵

Some of the dissertation's results did not reflect previous work. The study, for example, did not find as many personal or emotional tags as previous tagging studies have, perhaps indicating participants considered other's use of the tagged object rather

⁴⁰² Jimerson, "Archives 101 in a 2.0 World: The Continuing Need for Parallel Systems."

⁴⁰³ Ibid, 305.

⁴⁰⁴ Kipp and Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices."

⁴⁰⁵ Guy and Tonkin, "Folksonomies."

than their own personal use.⁴⁰⁶ A longitudinal study of digital archival tags might still indicate additional personal connections or use of tagging. The findings did not include the malicious, promotional, or general spam-like tagging behavior noted by Koutrika et al.⁴⁰⁷ This could be due to the closed nature of the study.

Regarding social tagging within archives, the range of tag types and number of unique tag terms reinforces Yakel's case study of social tagging of the Hague City Archives.⁴⁰⁸ Additionally, the level and breadth of the description offered by the generated tags meets users' needs and desires as described by Allison-Bunnell, Yakel, and Hauck's previous research on helpful metadata elements and users' opinions of Web 2.0 tools within digital archives.⁴⁰⁹ The study addresses the users' reliability concerns through both the lack of incorrect tags, and the matching of tags with unselected metadata and query-log terms. The dissertation also addresses the Chapman's concerns regarding "the ability of the average Internet user to leave un-moderated content."⁴¹⁰ Although the data indicate concerns are not necessary, the onus must be changing the users' perception of tags through outreach, and increasing the number of tags they see within digital archives.

5.2 Practical Implications

While the theoretical implications focus on previous studies and postmodernism, the dissertation study has broad practical implications. Data from the dissertation study indicates the benefits of including both expert and novice tags within a minimally

⁴⁰⁶ Kipp, "@toread and Cool: Subjective, Affective and Associative Factors in Tagging," and Golder and Huberman, "Usage Patterns of Collaborative Tagging Systems."

⁴⁰⁷ Koutrika et al., "Combating Spam in Tagging Systems."

⁴⁰⁸ Yakel, "Inviting the User into the Virtual Archives."

⁴⁰⁹ Allison-Bunnell, Yakel, and Hauck, "Researchers at Work."

⁴¹⁰ Chapman, "Observing Users," 24.

processed digital archive. Additionally, the findings of each research question and its associated sub-questions provide related specific practical implications. These individual implications are discussed in order of their associated research questions following the shared practical implications.

The largest implications of the dissertation's findings relate specifically to the application of tags within a minimally processed digital archive. The researcher posits in the dissertation's introduction, using prior domain knowledge as an indicator for tagging quality; specifically, restricting tagging to expert users. While the data analysis demonstrates a difference between expert and novice participants' tags, the categorical association is weak at best. In general, experts provided more content summary and contextualization tags by approaching tagging with a broader perspective than did novices. This is not suggesting novice users' tags are necessarily of lesser quality, however. While novice users did not produce as many content-summary tags, they were more adept at the subject tags, identifying persons, places, objects, and time periods within the photographs and documents.

The lack of large variations between experts and novices indicate negative results for the dissertation. The suggested approach of using domain knowledge as a quality assurance mechanism will not, according to the data, work effectively. Although disappointing at first glance, these results provide significant practical implications since the data refute many previous concerns regarding the application and use of tagging. The very low rate of incorrect tags (0.29% overall) should assuage critics' fears of tagging producing a gaggle of useless access points. Overall, the data demonstrate nothing positive about only including experts' tags. Rather, the exclusion of novice (and

intermediate) tags merely eliminates additional descriptions, interpretations, and ultimately, access points which would pair with similar users' search terms. As such, the author suggests the inclusion of both expert and novice tags within minimally processed digital collections.

The benefits of including both expert and novice tags is more clearly seen through the comparison with unselected metadata and query terms. The proportion of unselected metadata and query terms matching expert tags was higher than that matching novice tags. The combination of experts and novices, however, provided an even higher percentage thereby demonstrating the strength of incorporating both sets of tags into a collection. Additionally, since the study did not include intermediate users' tags (as is discussed later), the combination of all three might be even higher.

Rather than implying that one domain group should be trusted more than another, the results merely imply each grouping has different qualities, each serving differing purposes. If a collection prefers more content-summary tags, it should consider restricting tagging to expert users. A different mechanism for assessing domain knowledge might be considered, however, since the creation of a different domain-specific test for each collection would quickly become cumbersome. On the other hand, if a repository desires a broader range of access points to their minimally processed digital collections, they should not restrict the tagging based solely on prior domain knowledge.

The findings regarding incorrect tags and replication of metadata provide general tagging implications through the coding analysis' inclusion of both as major categories of tags. A major tagging concern from previous studies is the potential (or likelihood) of

incorrect tags. The dissertation project addressed this concern through including incorrect tags within its coding analysis, and found it to be the least occurring category throughout formats and domain groups, with only twenty-seven occurrences of incorrect tags out of 9,278 tags (0.29%). Similar to the replication of the metadata problem, the lack of incorrect tags within the dissertation analysis reaffirms previous findings, but at slightly lower rates.⁴¹¹ The influence of tagging conditions, specifically the limited number of taggers and non-natural development of tags could explain the lower level; however, the general replication of previous findings indicates a need for removal of incorrect tags as a primary concern within digital collections.

The coding scheme also addressed the issue of metadata replication and the analysis found 18.47% of all generated tags replicated the minimal metadata provided to participants. Jeong's two previous studies on YouTube tags both found a high degree of metadata replication among tags, with roughly half of the YouTube tags sampled matching previously used words in the title and/or description of the videos.⁴¹² In this case, the lack of detailed descriptions and titles might have reduced the proportion of metadata replication. Despite its reduction, metadata replication remains a concern and appeared in both domain groupings, suggesting a likely ongoing issue with tagging in general.

The participants' opinions regarding tagging presented several improvement suggestions, specifically, indicating that users desire more concrete instructions for tag creation (e.g., tutorials, guidelines, etc.). Repositories could specifically address the

⁴¹¹ Benoit III, "Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions."

⁴¹² Wooseob Jeong, "Does Tagging Really Work?" and Wooseob Jeong, "Is Tagging Effective? – Overlapping Ratios with Other Metadata Fields."

metadata replication problem through directing potential taggers not to create such tags. Of course, some taggers will not read/listen to directives, and will continue replicating metadata but, the proportions would be greatly reduced. The tags created for the dissertation study did not develop naturally, and many digital collections currently allow user tagging, but do not generate much tagging interest. Social tagging is not, and cannot be a “if we built it, they will come,” system. If repositories are truly interested in incorporating tagging into their collection plans, they need to reach out and provide better tagging mechanisms and support. Participants indicated a desire and willingness to tag collections; however, they are often unclear about what to tag or how to create tags. Repositories should consider including specific tagging instructions, which could also indicate the types of tags they would prefer.

Another concern is the usability of tagging systems. Many content management systems, such as CONTENTdm, added tagging and/or commenting tools to their software as an add-on during a version upgrade. Since the systems were not originally designed with user-generated content in mind, they are often clunky adaptations. Archives should work with vendors to create new, more intuitively designed products that allow for additional features and tagging controls, such as: suggested tags, spellchecking, abuse reporting, and tagger management tools (analysis of taggers by system administrators). Likewise, if an archive remains concerned over tagging consistency and terms, a better system could provide users with an approved list of tags to select from (a tagging-controlled vocabulary)—although this would negate the openness of tags and their postmodern potential.

In addition to the postmodern implications of tagging a minimally processed collection, another anticipated benefit was the potential for tags to replicate the unselected portions of traditional item-level metadata. The findings do not indicate a high level of replication of unselected metadata from either experts or novices. Even the combination of experts and novices did not produce more than 57% replication. This suggests the integration of tagging and minimal processing cannot completely replace the tradition item-level description/metadata of digital archives. In practical terms repositories considering allowing user tagging must be clear with their expectations and understand that tagging results in a different type of description.

Although the tags do not replicate the unselected metadata, they do serve as access points to the collection. Similar to previous points, the experts' tags again scored higher than novices', with the combination of both groups exceeding the individual groupings. A comparison of the proportion of *March on Milwaukee* query terms that match generated tags with those matching the unselected metadata shows a similar level (22.37% for tags, 24.74% for traditional metadata). This suggests the lack of matching unselected metadata is not as important when considering the terms users actually use for searching of the collection. In this case, the tags provide similar access to collection as that provided by their traditional metadata counterpart. Additionally, while the metadata terms in a collection are static, the number of unique tags would likely grow over time, thereby increasingly the likelihood of query terms matching tags to overtake the full metadata rates.

The study's findings provide practical implications for metadata creation, specifically by increasing the quality and breadth of metadata in a collection. Participants

created many tags which matched the real-world user query terms but did not match the unselected metadata. This implies users are searching for terms not included within the standard metadata corpus. Although users will always search for terms not found within a collection, the matching tags indicate the need to increase access points to the collection to best serve users' searching behavior. The documents in particular would benefit from additional content-driven or transcription-like metadata since those types of tags comprised the largest portion of the additional tags matching the query-log terms.

As noted in section 4.3, the real world metadata for the documents did not include the description Dublin Core element thereby leaving a significant deficiency within the item-level metadata. The tags matching query terms but not the unselected metadata would fill the description element well. A repository could use tag and query analyses to identify metadata gaps in both minimally processed and traditionally processed collections and develop new targeted strategies for filling the gaps.

Finally, the dissertation results suggest several practical recommendations for archival practitioners interested in social tagging. First, and foremost, social tags are value additive; that is to say, the inclusion of social tags increases access points, provides broader interpretations of the digital objects, and does not clutter the metadata with a swath of incorrect terminology. Archivists, therefore, should approach tagging with confidence towards its benefits rather than with unwarranted hesitation or fearfulness. Secondly, archivists should provide some basic instruction or prompting to direct the creation of tags and the types of tags they desire. This can be accomplished through a description on the home page, a well worded email, or even through prompting users viewing items themselves. Finally, while tags may not entirely replace item-level

metadata they do provide enough coverage that questions the need for the labor-intensive practice of item-level description.

5.3 Methodological Implications

The dissertation relied on a mixed-method, quasi-experiment design in addressing the research questions. This approach, and the associated methodological steps taken, provides some limited methodological implications for future research. First and foremost, the use of experimental or quasi-experimental designs in archival research remains limited. The dissertation study's design offers a model for future archival user research into innovative tools and solutions to archival issues; a need noted by Conway's framework of archival user studies.⁴¹³ Additionally, the dissertation successfully relied on a domain knowledge assessment mechanism for placing participants into appropriate groupings rather than self-selection or a post hoc placement based on the user's results. This innovative approach offers an additional model for future research designs, resulting in improved quality of future findings.

The dissertation study's design relied entirely on a Web-based structure, from the recruiting of participants through the post-questionnaire. Multiple linked survey mechanisms built in Qualtrics routed participants through the study's stages via utilization of Qualtrics' built-in skip/display logic and quota systems. The remote nature of the design allowed the researcher limited interaction with the participants, thereby limiting any influence on the results. More importantly, the online structure removed recruitment physical location barriers, thereby increasing the geographic variety of the participants. Similar to the assessment mechanism, eliminating physical requirements

⁴¹³ Conway, "Facts and Frameworks: An Approach to Studying the Users of Archives."

from a study can open the door for a wide range of additional research, such as comparative research. Finally, the design incorporated both real users and a realistic setting through directed recruitment and the hosting of the sample collection within CONTENTdm.

5.3 Limitations

All research requires difficult decisions in planning and execution, inevitably resulting in specific data and result limitations. The dissertation project is not an exception. First and foremost, through designing a quasi-experimental study, the dissertation focused on prior domain knowledge as the primary independent variable. The pre-questionnaire knowledge assessment placed all interested participants into one of three groups: experts, intermediates, and novices. Participants in the intermediate group were immediately dismissed from the study, isolating the domain knowledge extremes (experts and novices) for a better comparison. Although the decision served its intended purpose, additional factors (e.g., number of tags produced, time spent tagging, etc.) could not be explored since a third of the population was already removed.

Similarly, the tagging conditions of the study limit the generalizability of the findings, although some features minimized this limitation. Requiring each participant to provide a minimum number of tags meant the tags did not develop as naturally as would be viewed in a longitudinal study. The minimum requirement was necessary to provide a sufficient mass of tags for coding analysis. The sample collection only provided access to the taggable items in the collection and did not include a full version of the archival collection. The sample collection preserved the contextual information between the photographs and documents through the structural arrangement and shared metadata;

however, access to other materials in the collection focused participants on the 30 specific items.

The use of separate instances of the sample collection provided another limitation. The separation of expert and novice collection was necessary to isolate the tags generated by each group, while a real-world application would have all users interacting with the same collection. Hosting the sample collection instances in the CONTENTdm interface provided a more realistic participant experience, although some of the controls required a slight learning curve. The participant instructions included a tutorial video to best negate this issue.

Additional conditions of the study limited the findings' generalizability. The dissertation used a total of sixty participants, and a sample collection of 30 records. Both groups are a relatively small sample size, and additional future studies would be needed to replicate the findings to increase their generalizability. The specific collection selected, *March on Milwaukee*, could have some unknown impact on the findings. As of now, the study can only reliably state its findings are true for a very similar collection. Future studies should examine tagging within different subject matter collections for a fuller understanding of tagging behavior.

Finally, the dissertation study focused on two format types: photographs and documents. Inclusion of additional archival formats, such as audio, moving image, or cartographic, could result in different data and conclusions. The choice to focus on photographs and documents was made based on the current popularity of these formats within digital archives. Future research will address additional formatted materials.

CHAPTER SIX: CONCLUSION

During the past decade, minimal processing quickly rose as a potential solution for addressing archival backlogs and the increased amounts of incoming collections. Although initially focused on the archival processes of arrangement and description, minimal processing expanded to all segments of the archival endeavor. Its use within digital collections increases the amount of digital records available to users; however, the limited amount of metadata combined with the lack of unique item-level metadata limits access to the materials.

Nestled within postmodernism, the participatory archives movement offers a solution through user-generated social tags. Previous research, however, indicates a high degree of mistrust regarding quality and consistency of tags by both users and archivists. The dissertation project explores the possibility of mediating tags (and thereby maintaining quality) through inclusion of only expert domain users' tags in a minimally processed digital archive. Focusing on three main research questions, the quasi-experimental design study highlighted the difference between tags created by novice and expert users, compared the tags with unselected metadata—the item-level metadata from a traditionally processed collection, and compared the tags with real-world user query terms. Figure 7.1 summarizes the dissertation's research questions, findings, and implications.

Sixty participants divided into two groups created a total of 9,278 tags, of which 1,463 were unique. On the whole, both experts and novices created more tags for the document than the photographs. Novices generated more tags for each format grouping;

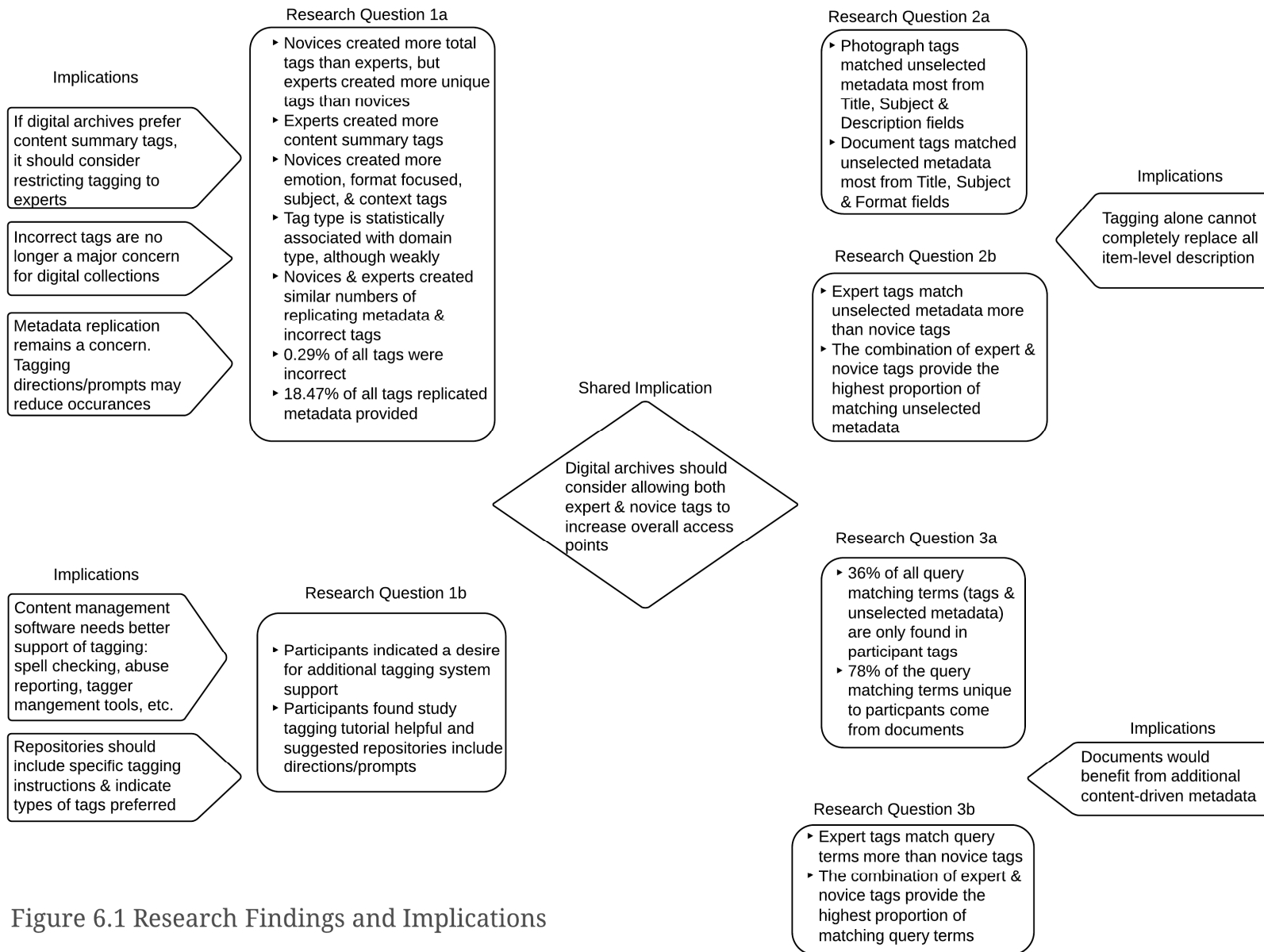


Figure 6.1 Research Findings and Implications

however, experts created more unique tags. The differences between the number of tags generated by experts and novices was not statistically significant. A coding analysis of the tags identified seven major categories and two subcategories of tags: replication of metadata, format focused, identification (with subcategories of general and specific), description, context, emotion, and incorrect. Statistical analysis found a weak association between domain group (expert, novice) and the tag categories, with experts creating more descriptive and context tags, while novices created more identification tags. The association remained when analyzing the different format groups as well.

The comparison of expert and novice tags with the unselected metadata found a low proportion of tags matched the unselected metadata for both photographs in the aggregate and novices. Although the levels were low, experts' tags matched at a higher rate than novices' in both photographs and documents. Statistical analysis found a weak association between domain knowledge group and the proportion of photographic tags matching the unselected metadata. The analysis of the document tags did not find a statistically significant association.

Finally, the comparison of expert and novice tags with existing user query terms reflected the unselected metadata comparison, with the proportion of query terms matching experts' tags at a higher level than novices'. Statistical tests found another weak association between domain knowledge group and proportion of query terms matching tags. The analysis was similar when using *March on Milwaukee*-specific query terms and cross-collection query terms.

Overall, the dissertation found a difference exists between the expert and novice tags; however, the differences in all aspects are minimal. Although a minimally processed archive could rely on prior domain knowledge level as a mediation mechanism, the resulting tags would not provide a well-rounded multi-perspective interpretation of the records. The benefits of including tags from both groups are clearly seen through the results from research questions two and three since the proportion of unselected metadata or query terms matching tags was highest when combining the tags from experts and novices.

6.1 Future Research

The results, implications, and limitations of the dissertation naturally lead toward continued and future research themes and applications. Specifically addressing the limitations of excluding the intermediate users from the study, additional research should focus on exploring additional alternative factors that may produce greater differences between groups. These factors include, but are not limited to, the number of tags generated per user (focusing on the influence of so-called super taggers), time spent tagging, tagger's age, and the division of researchers and non-researchers. Similarly, additional future studies should include additional archival formats to better compare the tagging efficiency and efficacy. Formats, such as audio and moving images, may produce different results since they would require increased attention from the participants (due to the nature of the formats themselves).

The dissertation used a non-natural tag development technique within its quasi-experimental design. This required particular sacrifices, which should be the focus of future studies. A longitudinal study could analyze the natural development of tags within

a larger collection and could also integrate the participants into one collection (rather than the separate collections of this study). Although the results would not share the experimental nature of the dissertation, the longitudinal version's results would be more directly applicable for real-world digital archives.

The results of research questions two and three focused on the comparison of tags with the unselected metadata and real-world user query terms. Future research should further explore the information retrieval effectiveness of social tags through pure experimental designs resulting in concrete empirical data. Addressing the results of this study specifically, a future experimental design with double-blind random assignment could compare the IR effectiveness between three conditions: a full metadata collection; a minimal metadata collection; and a minimal metadata collection with tags. The experiment could focus on both the participant experience, and their success at locating known items or subject searches for each collection set. Additional future studies could also analyze a larger query-log set and compare the query terms with a fully tagged collection.

Finally, each group as it currently stands could be further delineated into subgroupings. The monetary rewards, for example, could go beyond direct payment to taggers; it could also include discounted memberships, free memberships, photocopies, photographic prints, etc.

6.2 Future Directions

Although the dissertation's findings could not entirely support the use of prior domain knowledge as a quality assurance mechanism for tags, the results provide

optimism for the use of all tags regardless of the user's domain knowledge; essentially rejecting the need for quality assurance mechanisms entirely. Additionally, the findings should further ease archivists' concern over incorrect tags and the need for continuous, active monitoring of a tagging environment. Without oversight, tags can and will develop an increased level of digital material description and access points over time, and by not limited the tagging to specific users, archives will continue striving for inclusiveness of opinions and perspectives rather than return to the exclusionary past.

In a broader sense, the dissertation's findings and recommendations strengthen the postmodern approach through not excluding voices from the archives while limiting the inherent bias of the processing archivist. If the findings had supported using domain knowledge as a quality assurance mechanism, the postmodern approach would not be adequately met since the archives would simple switch preference for one bias voice (the archivist) toward a different exclusionary voice (the expert). Instead, an nearly ideal postmodern condition emerges, and can be used toward building a more inclusive archival community.

Minimal processing and MPLP addressed hidden collections and backlogs through prioritizing collection access as a whole rather than individual record access. Rather than interpreting the increased use of MPLP as a trend toward limiting access points, archivists should grasp the opportunity to further connect with their users and communities by engaging the service of taggers and diversifying the archival voice in the process. Asking for users' assistance in daunting task of description will also increase the visibility of archives in society, thereby raising their inherent value. Of course, this also requires directed appeals rather than simply turning on the tagging function within a

collection. Archives and archivists must demonstrate the benefits of creating tags through demonstrations, tutorials, and instructional materials. Further appeals could be approached from a wide range of methods including tagging games, offering intrinsic or non-intrinsic rewards, and other yet tested mechanisms. Archivists have always proved resourceful, and this provides yet another opportunity for innovation and testing.

Overall, the trend of archival practice must be toward user engagement and interaction rather than away from it. Just as the business world, educators, and government agencies rely on crowdsourcing to help fill missing gaps of information, so to must archives and archivists. Currently, the lack of access points is an irritation for many users, but with the continuous and unrelenting increase of digital materials, it will soon make some archives unusable. This will result in a similar problem that MPLP was originally designed to solve, the backlog problem and collections without access. Only through embracing combinations of automatic metadata creation, minimal metadata, and user-generated tags will archives remain viable.

Bibliography

- Abraham, Terry, Stephen E. Balzarini, and Anne Frantilla. "What is Backlog is Prologue: A Measurement of Archival Processing." *American Archivist* 48, no. 1 (1985): 31–44.
- Agosti, Maristella, and Nicola Ferro. "A Formal Model of Annotations of Digital Content." *ACM Transactions on Information Systems* 26, no. 1 (November 2007): 3:1-3:57.
- Agosti, Maristella, Nicola Ferro, Emanuele Panizzi, and Rosa Trinchese. "Annotation As a Support to User Interaction for Content Enhancement in Digital Libraries." In *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York: ACM, 2006.
- Akmon, Dharma. "Only with Your Permission: How Rights Holders Respond (or Don't Respond) to Requests to Display Archival Materials Online." *Archival Science* 10, no. 1 (2010): 45–64.
- Allison-Bunnell, Jodi, Elizabeth Yakel, and Janet Hauck. "Researchers at Work: Assessing Needs for Content and Presentation of Archival Materials." *Journal of Archival Organization* 9, no. 2 (2011): 67–104.
- Altman, Burt and John R. Nemmers. "The Usability of On-Line Archival Resources: The Polaris Project Finding Aid." *American Archivist* 64, no. 1 (2001): 121–131.

Ames, Morgan and Mor Naaman. "Why We Tag: Motivations for Annotation in Mobile and Online Media." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, 2007.

Anderson, Karen, Jeannette Bastian, Ross Harvey, Terry Plum, and Göran Samuelsson. "Teaching to Trust: How a Virtual Archives and Preservation Curriculum Laboratory Creates a Global Education Community?" *Archival Science* 11, no. 3–4 (2011): 349–372.

Anderson, Scott R. and Robert B. Allen. "Envisioning the Archival Commons." *American Archivist* 72, no. 2 (2009): 383–400.

Andreano, Kevin. "The Missing Link: Content Indexing, User-Created Metadata, and Improving Scholarly Access to Moving Image Archives." *The Moving Image* 7, no. 2 (2007): 82–99.

Astle, Peter J., and Adrienne Muir. "Digitization and Preservation in Public Libraries and Archives." *Journal of Librarianship and Information Science* 34, no. 2 (2002): 67–79.

Bak, Greg. "Continuous Classification: Capturing Dynamic Relationships among Information Resources." *Archival Science* 12, no. 3 (2012): 287–318.

Balkestein, Marjan, and Heiko Tjalsma. "The ADA Approach: Retro-Archiving Data in an Academic Environment." *Archival Science* 7, no. 1 (2007): 89–105.

Baxter, Terry D. "Going to See the Elephant: Archives, Diversity, and the Social Web."

In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 274-303. Chicago: Society of American Archivists, 2011.

Bearman, David and Jennifer Trant. "Social Terminology Enhancement through

Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections." *D-Lib Magazine* 11, no. 9 (2005).

<http://www.dlib.org/dlib/september05/bearman/09bearman.html>.

Bearman, David and Margaret Hedstrom. "Commentary Reinventing Archives for

Electronic Records: Alternative Service Delivery Options." In *Electronic Records Management Program Strategies*, edited by Margaret Hedstrom, 82-98.

Pittsburgh, PA: Archives & Museum Informatics, 1993.

Bearman, David. "Archival Principles and the Electronic Office." In *Electronic*

Evidence: Strategies for Managing Records in Contemporary Organizations, edited by David Bearman, 145-175. Pittsburgh, PA: Archives & Museum Informatics, 1994.

———. "Information Technology Standards and Archives." In *Electronic Evidence:*

Strategies for Managing Records in Contemporary Organizations, edited by David Bearman, 210-221. Pittsburgh, PA: Archives & Museum Informatics, 1994.

———. "Interactive and Hypermedia in Museums." In *Hypermedia & Interactivity in*

Museums, Proceedings of an International Conference, 1-6. Pittsburgh, PA: Archives & Museum Informatics, 1991.

- . “New Models for Management of Electronic Records.” In *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, edited by David Bearman, vol. 2, 278-292. Pittsburgh, PA: Archives & Museum Informatics, 1994.
- . “Reality and Chimeras in the Preservation of Electronic Records.” *D-Lib Magazine* 5, no. 4 (April 1999): 1–5.
- . “Virtual Archives.” Presented at the ICA Meeting, Beijing, China, 1996.
<http://web.archive.org/web/19990427133904/http://www.lis.pitt.edu/~nhprc/prog6.html>.
- Becker, Devin and Collier Nogues. “Saving-Over, Over-Saving, and the Future Mess of Writers’ Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers.” *American Archivist* 75, no. 2 (2012): 482–513.
- Benoit, III, Edward. “Digital Librarians’ Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations.” 2012, Manuscript in Preparation.
- . “Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions.” 2012, Manuscript in Preparation.
- Benson, Allen C. “The Archival Photograph and Its Meaning: Formalisms for Modeling Images.” *Journal of Archival Organization* 7, no. 4 (2009): 148-187.
- Borgman, C. L. “The User’s Mental Model of an Information Retrieval System: An Experiment on a Prototype Online Catalog.” *International Journal of Man-Machine Studies* 24, no. 1 (1986): 47–64.

- Bowen Maier, Shannon. "MPLP and the Catalog Record as a Finding Aid." *Journal of Archival Organization* 9, no. 1 (2011): 32–44.
- Boyer, Deborah, Robert Cheetham, and Mary Johnson. "Using GIS to Manage Philadelphia's Archival Photographs." *American Archivist* 74, no. 2 (2011): 652–663.
- Bültmann, Barbara, Rachel Hardy, Adrienne Muir, and Clara Wictor. "Digitized Content in the UK Research Library and Archives Sector." *Journal of Librarianship and Information Science* 38, no. 2 (2006): 105–122.
- Byrne, Gillian. "A Statistical Primer: Understanding Descriptive and Inferential Statistics." *Evidence Based Library and Information Practice* 2, no. 1 (2007): 32–47.
- Campbell, Donald T. and Julian C Stanley. *Experimental and Quasi-Experimental Design for Research*. Chicago: Rand-McNally, 1963.
- Cantor, David, Barbara C. O'Hare, and Kathleen S. O'Connor. "The Use of Monetary Incentives to Reduce Nonresponse in Random Digit Dial Telephone Surveys." In *Advances in Telephone Survey Methodology*, edited by James M. Lepkowski et al., 471-498. Hoboken, NJ: John Wiley & Sons, 2008.
- Carroll, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. "A Comprehensive Approach to Born-Digital Archives." *Archivaria* 72, no. Fall (2011): 61–92.
- Caswell, Michelle. "On Archival Pluralism: What Religious Pluralism (and Its Critics) Can Teach Us about Archives." *Archival Science* 13, no. 4 (2013): 273–292.

- Cattuto, Ciro, Andrea Baldassarri, Vito D.P. Servedio, and Vittorio Loreto. "Emergent Community Structure in Social Tagging Systems." *Advances in Complex Systems*, 11, no. 4 (2008): 597–608.
- Chan, Sebastian. "Tagging and Searching--Serendipity and Museum Collection Databases." In *Museums and the Web 2007: Proceedings*, 2007.
<http://www.archimuse.com/mw2007/papers/chan/chan.html>.
- Chandler, Robin L. "Building Digital Collections at the OAC." *Journal of Archival Organization* 1, no. 1 (2002): 93–103.
- Chapman, Joyce Celeste "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids." *Journal of Archival Organization* 8, no. 1 (2010): 4–30.
- Chen, Chufeng, Michael Oakes, and John Tait. "A Location Annotation System for Personal Photos." In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '06*, 2006. <http://doi.acm.org/10.1145/1148170.1148339>.
- Chi, Michelene T. H., Robert Glaser, and Marshall J. Farr, eds. *The Nature of Expertise*. Hillsdale, NJ: Erlbaum, 1988.
- Christen, Kimberly. "Opening Archives: Respectful Repatriation." *American Archivist* 74, no. 1 (2011): 185–210.
- Christian, Michele and Tanya Zanish-Belcher. "'Broadcast Yourself': Putting Iowa State University's History on YouTube." In *A Different Kind of Web: New Connections*

Between Archives and Our Users, edited by Kate Theimer, 33-41. Chicago, IL: Society of American Archivists, 2011.

Chung, EunKyung and JungWon Yoon. "Categorical and Specificity Differences between User-Supplied Tags and Search Query Terms for Images: An Analysis of Flickr Tags and Web Image Search Queries." *Information Research* 14, no. 3 (2009). <http://www.informationr.net/ir/14-3/paper408.html>.

Coleman, Sterling. "The Archival and Library Viewpoints of a Collection in a Digital Environment: Is There Any Room for Compromise?" *Journal of Archival Organization* 2, no. 1 (2004): 103-115.

Combs, Michele. "Wikipedia as an Access Point for Manuscript Collections." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 139-147. Chicago: Society of American Archivists, 2011.

Conway, Paul. "Archival Quality and Long-Term Preservation: A Research Framework for Validating the Usefulness of Digital Surrogates." *Archival Science* 11, no. 3 (2011): 293-309.

———. "Facts and Frameworks: An Approach to Studying the Users of Archives." *American Archivist* 49, no. 4 (1986): 393-407.

———. *Preservation in the Digital World*. Washington, DC: Commission on Preservation and Access, 1996.

Cook, Terry. "Archival Science and Postmodernism: New Formulations for Old Concepts." *Archival Science* 1, no. 1 (2001): 3-24.

- Corbin, Juliet M. and Anselm Strauss. "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology* 13, no. 1 (1990): 3-21.
- Cox, Andrew, Paul Clough, and Stefan Siersdorfer. "Developing Metrics to Characterize Flickr Groups." *Journal of the American Society for Information Science and Technology* 62, no. 3 (2011): 493–506.
- Cox, Richard J. "Yours Ever (well, Maybe): Studies and Signposts in Letter Writing." *Archival Science* 10, no. 4 (2010): 373–388.
- Cox, Robert S. "Maximal Processing, Or, Archivist on a Pale Horse." *Journal of Archival Organization* 8, no. 2 (2010): 134–148.
- Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd edition. London: Sage, 2009.
- Crowe, Stephanie H. and Karen Spilman. "MPLP @ 5: More Access, Less Backlog?" *Journal of Archival Organization* 8, no. 2 (2010): 110–133.
- Crymble, Adam. "An Analysis of Twitter and Facebook Use by the Archival Community." *Archivaria* 70 (2010): 125–151.
- Cunningham, Adrian. "Digital Curation/Digital Archiving: A View from the National Archives of Australia." *American Archivist* 71, no. 2 (2008): 530-543.
- Davis, Susan E. "Electronic Records Planning in 'Collecting' Repositories." *American Archivist* 71, no. 1 (2008): 167–189.

Delort, Jean-Yves. "Automatically Characterizing Saliency Using Readers' Feedback."

Journal of Digital Information 10, no. 1 (2009).

<http://journals.tdl.org/jodi/index.php/jodi/article/view/268>.

DeRidder, Jody, Amanda Presnell, and Kevin Walker. "Leveraging Encoded Archival Description for Access to Digital Content: A Cost and Usability Analysis."

American Archivist 75, no. 1 (April 1, 2012): 143–170.

Derrida, Jacques. *Archive Fever: A Freudian Impression*. Chicago and London:

University of Chicago Press.

Desnoyers, Megan. "When Is It Processed?" *Midwestern Archivist* 7, no. 1 (1982): 5–23.

DeVellis, Robert F. *Scale Development: Theory and Applications*. Newbury Park, CA:

Sage, 1991.

Dickson, Maggie. "Due Diligence, Futile Effort: Copyright and the Digitization of the

Thomas E. Watson Papers." *American Archivist* 73, no. 2 (2010): 626–636.

Donaldson, Devan Ray and Elizabeth Yakel. "Secondary Adoption of Technology

Standards: The Case of PREMIS." *Archival Science* 13, no. 1 (2013): 55–83.

Doorn, Peter and Heiko Tjalsma. "Introduction: Archiving Research Data." *Archival*

Science 7, no. 1 (2007): 1–20.

Dryden, Jean. "Measuring Trust: Standards for Trusted Digital Repositories." *Journal of*

Archival Organization 9, no. 2 (2011): 127–130.

- Dryden, Jean. "Copyright Issues in the Selection of Archival Material for Internet Access." *Archival Science* 8, no. 2 (2008): 123–147.
- Dryden, Jean. "The Open Archival Information System Reference Model." *Journal of Archival Organization* 7, no. 4 (2009): 214–217.
- Duff, Wendy M. and Joan M. Cherry. "Archival Orientation for Undergraduate Students: An Exploratory Study of Impact." *American Archivist* 71, no. 2 (2008): 499–529.
- Duff, Wendy M., Amy Marshall, Carrie Limkilde, and Marlene van Ballegooie. "Digital Preservation Education: Educating or Networking?" *American Archivist* 69, no. 1 (2006): 188–212.
- Duff, Wendy M., Jean Dryden, Carrie Limkilde, Joan Cherry, and Ellie Bogomazova. "Archivists' Views of User-Based Evaluation: Benefits, Barriers, and Requirements." *American Archivist* 71, no. 1 (2008): 144–166.
- Duranti, Luciana and Randy Preston, eds. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008.
- Duranti, Luciana and Kenneth Thibodeau. "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES." *Archival Science* 6, no. 1 (2006): 13–68.
- Edmunson-Morton, Tiah. "Talking and Tagging: Using CONTENTdm and Flickr in the Oregon State University Archives." *The Interactive Archivist: Case Studies in*

Utilizing Web 2.0 to Improve the Archival Experience, June 19, 2009.

<http://interactivearchivist.archivists.org/case-studies/flickr-at-osu/>.

Erway, Ricky and Jennifer Schaffner. *Shifting Gears: Gearing up to Get Into the Flow*.

Dublin, OH: OCLC Programs and Research, 2007.

<http://www.oclc.org/programs/publications/reports/2007-02.pdf>.

Evans, Max J. "Archives of the People, by the People, for the People." *American*

Archivist 70, no. 2 (2007): 387-400.

Evans, Gwen and Susannah Cleveland. "Moody Blues: The Social Web, Tagging, and

Nontextual Discovery Tools for Music." *Music Reference Services Quarterly* 11,

no. 3-4 (2008): 177-201.

Evans, Joanne, Sue McKemmish, and Karuna Bhoday. "Create Once, Use Many Times:

The Clever Use of Recordkeeping Metadata for Multiple Archival Purposes."

Archival Science 5, no. 1 (2005): 17-42.

Eveleigh, Alexandra. "Welcoming the World: An Exploration of Participatory Archives."

Presented at the *International Conference on Archives*. Brisbane, Australia, 2012.

<http://www.gosbook.ru/system/files/documents/2012/11/13/ica12Final00128.pdf>.

Fernandes, Marco, Miguel Aljo, Joaquim Arnaldo Martins, Joaquim Sousa Pinto, and

Pedro Almedia. "Web Annotation System Based on Web Services." In

Proceedings of the International Conference on Next Generation Web Services

Practices, 2005. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1592399.

- Fisher, Barbara. "Byproducts of Computer Processing." *American Archivist* 32, no. 3 (1969): 215–223.
- Flinn, Andrew. "An Attack on Professionalism and Scholarship? Democratising Archives and the Production of Knowledge." *Ariadne* 62 (2010).
<http://www.ariadne.ac.uk/issue62/flinn>.
- Ford, Nigel, David Miller, and Nicola Moss. "The Role of Individual Differences in Internet Searching: An Empirical Study." *Journal of the American Society for Information Science and Technology* 52, no. 12 (2001): 1049–1066.
- . "Web Search Strategies and Human Individual Differences: A Combined Analysis." *Journal of the American Society for Information Science and Technology* 56, no. 7 (2005): 757–764.
- Foster, Anne L. "Minimum Standards Processing and Photograph Collections." *Archival Issues* 30, no. 2 (2006): 107–118.
- Fox, Robert. "Cataloging for the Masses." *OCLC Systems & Services: International Digital Library Perspectives* 22, no. 3 (2006): 166–172.
- Fredriksson, Berndt. "Postmodernistic Archival Science — Rethinking the Methodology of a Science." *Archival Science* 3, no. 2 (2003): 177–197.
- French, Christopher. "Computerizing London's Eighteenth-Century Maritime Activity." *Archives: Journal of BRA* 22, no. 97 (1997): 130–140.
- Frusciano, Thomas J. "'The Beat Goes On': MPLP, RDA, Digitization, and Archivist as Historian." *Journal of Archival Organization* 8, no. 3–4 (2010): 169–173.

- Fu, Xin, Tom Ciszek, Gary Marchionini, and Paul Solomon. "Annotating the Web: An Exploratory Study of Web Users' Needs for Personal Annotation Tools." In *Proceedings of the American Society for Information Science and Technology*, vol. 42, 2005. <http://onlinelibrary.wiley.com/doi/10.1002/meet.14504201151/full>.
- Furner, Jonathan. "User Tagging of Library Resources: Toward a Framework for System Evaluation." Presented at the *World Library and Information Congress: 73rd IFLA General Conference and Council*. Durban, South Africa, 2007. <http://ifla.queenslibrary.org/iv/ifla73/papers/157-Furner-en.pdf>.
- Gahan, Paul. "Social Networking, the Swindon Collection." *Multimedia Information and Technology* 36, no. 4 (2010): 25–27.
- Galloway, Patricia. "Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice." *Archivaria* 72 (2011): 169–196.
- Garvin, Peggy. "Photostreams to the People: The Commons on Flickr." *Searcher* 17, no. 8 (2009): 45–49.
- Gavrilis, Dimitris, Constantia Kakali, and Christos Papatheodorou. "Enhancing Library Services with Web 2.0 Functionalities." In *Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science* 5173. Springer Berlin Heidelberg, 2008, 148–159. http://link.springer.com/chapter/10.1007/978-3-540-87599-4_16.

- Gerencser, James. "New Tools Equal New Opportunities: Using Social Media to Achieve Archival Management Goals." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 159-179. Chicago: Society of American Archivists, 2011.
- Gilliland-Swetland, Anne J. "An Exploration of K-12 User Needs for Digital Primary Source Materials." *American Archivist* 61, no. 1 (1998): 136-157.
- . "Testing Our Truths: Delineating the Parameters of the Authentic Archival Electronic Record." *American Archivist* 65, no. 2 (2002): 196-215.
- Goh, Dion H. and Alton Y. K. Chua. "A Study of Web 2.0 Applications in Library Websites." *Library & Information Science Research* 32, no. 3 (2010): 203-211.
- Golder, Scott A. and Bernardo A. Huberman. "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science* 32, no. 2 (2006): 198-208.
- Göritz, Anja S. "Contingent versus Unconditional Incentives in WWW-Studies." *Advances in Methodology & Statistics* 2, no. 1 (2005): 1-14.
- Göritz, Anja S. "Incentives in Web Studies: Methodological Issues and Review." *International Journal of Internet Science* 1, no. 1 (2006): 58-70.
- Gorzalski, Matt. "Minimal Processing: Its Context and Influence in the Archival Community." *Journal of Archival Organization* 6, no. 3 (2008): 186-200.
- Gracy, Karen. "Distribution and Consumption Patterns of Archival Moving Images in Online Environments." *American Archivist* 75, no. 2 (2012): 422-455.

- Grannum, Guy. "Harnessing User Knowledge: The National Archives' Your Archives Wiki." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 116-127. Chicago: Society of American Archivists, 2011.
- Greenberg, Jane. "The Applicability of Natural Language Processing (NLP) to Archival Properties and Objectives." *American Archivist* 61, no. 2 (1998): 400-425.
- Greene, Mark A. and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68, no. 2 (2005): 208-263.
- Greene, Mark A. "Doing Less Before It's Done Unto You: Reshaping Workflows for Efficiency Before the Wolf Is at the Door." *RBM: A Journal of Rare Books, Manuscripts and Cultural Heritage* 12, no. 2 (2011): 92-103.
- . "MPLP: It's Not Just for Processing Anymore." *American Archivist* 73, no. 1 (2010): 175-203.
- Grover, Wayne C. "Recent Developments in Federal Archival Activities." *American Archivist* 14, no. 1 (1951): 3-12.
- Gupta, Manish, Rui Li, Zhijun Yin, and Jiawei Han. "Survey on Social Tagging Techniques." *ACM SIGKDD Explorations Newsletter* 12, no. 1 (2010): 58-72.
- Guy, Marieke and Emma Tonkin. "Folksonomies: Tidying up Tags?" *D-Lib Magazine* 12, no. 1 (2006). <http://www.dlib.org/dlib/january06/guy/01guy.html>.
- Haller, Uli. "Variations in the Processing Rates on the Magnuson and Jackson Senatorial Papers." *American Archivist* 50, no. 1 (1987): 100-109.

- Ham, F. Gerald, Frank Boles, Gregory S. Hunter, and James M. O'Toole. "Is the Past Still Prologue?: History and Archival Education." *American Archivist* 56, no. 4 (1993): 718–729.
- Hammond, Tony, Timo Hannay, Ben Lund, and Joanna Scott. "Social Bookmarking Tools (I): A General Review." *D-Lib Magazine* 11, no. 4 (2005).
<http://www.dlib.org/dlib/april05/hammond/04hammond.html>.
- Hank, Carolyn and Barbara M. Wildemuth. "Quasi-Experimental Studies." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 93-104. Westport, CT: Libraries Unlimited, 2009.
- Harris, Verne. "Claiming Less, Delivering More: A Critique of Positivist Formulations on Archives in South Africa." *Archivaria* 44 (1997): 132-141.
- Hasson, Dan and Bengt B. Arnetz. "Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements." *Global Journal of Health Education and Promotion* 8, no. 1 (2005): 178–192.
- Heald, Carolyn. "Is There Room for Archives in the Postmodern World?" *American Archivist* 59, no. 1 (1996): 88-101.
- Hedstrom, Margaret. "Archives, Memory, and Interfaces with the Past." *Archival Science* 2, no. 1–2 (2002): 21–43.
- . "Descriptive Practices for Electronic Records: Deciding What Is Essential and Imagining What Is Possible." *Archivaria* 1, no. 36 (1993): 53–63.

Hembrooke, Helene A., Laura A. Granka, Geraldine K. Gay, and Elizabeth D. Liddy.

“The Effects of Expertise and Feedback on Search Term Selection and Subsequent Learning.” *Journal of the American Society for Information Science and Technology* 56, no. 8 (2005): 861–871.

Henry, Gary T. *Practical Sampling*. Newbury Park, CA: Sage, 1990.

Hill, Janette R. and Michael J. Hannafin. “Cognitive Strategies and Learning from the World Wide Web.” *Educational Technology Research and Development* 45, no. 4 (1997): 37–64.

Hirsh, Sandrag. “How Do Children Find Information on Different Types of Tasks? Children’s Use of the Science Library Catalog.” *Library Trends* 45, no. 4 (1997): 725–745.

Hite, Richard W. and Daniel J. Linke. “Teaming Up with Technology: Team Processing.” *Midwestern Archivist* 15, no. 2 (1990): 91–98.

Hölscher, Christoph and Gerhard Strube. “Web Search Behavior of Internet Experts and Newbies.” *Proceedings of the 9th International World Wide Web Conference on Computer Networks: The International Journal of Computer and Telecommunications Networking* 33, no. 1 (2000): 337–346.

Holsti, Ole R. *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley, 1969.

Holz, Dayna. “Technologically Enhanced Archival Collections: Using the Buddy System.” *Journal of Archival Organization* 4, no. 1–2 (2007): 29–44.

- Hsieh, Hsiu-Fang and Sarah E. Shannon. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research* 15, no. 9 (2005): 1277-1288.
- Hunter, Jane. "Collaborative Semantic Tagging and Annotation Systems." *Annual Review of Information Science and Technology* 43, no. 1 (2009): 1-84.
- Huotari, Maija-Leena and Marjo Rita Valtonen. "Emerging Themes in Finnish *Archival Science* and Records Management Education." *Archival Science* 3, no. 2 (2003): 117-129.
- Huvila, Isto. "Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management." *Archival Science* 8, no. 1 (2008): 15-36.
- InterPARES Project. *InterPARES 3 Project*, 2012.
http://www.interpares.org/ip3/ip3_index.cfm.
- . *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, 2012. <http://www.interpares.org/book/index.cfm>.
- Jantz, Ronald and Michael Giarlo. "Digital Archiving and Preservation: Technologies and Processes for a Trusted Repository." *Journal of Archival Organization* 4, no. 1 (2006): 193-213.
- Jenkins, Christine, Cynthia L. Corritore, and Susan Wiedenbeck. "Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise." *IT & Society* 1, no. 3 (2003): 64-89.

Jeong, Wooseob. "Does Tagging Really Work?" In *Proceedings of the American Society for Information Science and Technology*, vol. 45, 2008.

<http://onlinelibrary.wiley.com/doi/10.1002/meet.2008.14504503124/full>.

———. "Is Tagging Effective? – Overlapping Ratios with Other Metadata Fields." *International Conference on Dublin Core and Metadata Applications* (2009): 31–39.

Jimerson, Randall C. "Archives 101 in a 2.0 World: The Continuing Need for Parallel Systems." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 304-333. Chicago: Society of American Archivists, 2011.

Johnson, Gregory P. "Quality or Quantity: Can Archivists Apply Minimal Processing to Electronic Records?" 2007. <http://www.ils.unc.edu/MSpapers/3267.pdf>.

Jones, Barbara M. "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in America's Research Libraries." *RBM: A Journal of Rare Books, Manuscripts and Cultural Heritage* 5, no. 2 (2004): 88–105.

Jones, Tricia. "Incidental Learning during Information Retrieval: A Hypertext Experiment." In *Computer Assisted Learning: Proceedings of the Second International Conference*, edited by Hermann Maurer, 235-253. Berlin: Springer-Verlag, 1989.

- Josephson, Bertha E. "How Can We Improve Our Historical Societies?" *American Archivist* 8, no. 3 (1945): 194-201.
- Kelly, Diane. "Methods for Evaluating Interactive Information Retrieval Systems with Users." *Foundations and Trends in Information Retrieval* 3, no. 1-2 (2009): 1-224.
- Ketelaar, Eric. "Being Digital in People's Archives." *Archives & Manuscripts* 31, no. 2 (2003): 8-22.
- . "Commentary on 'Archival Strategies': The Archival Image." *American Archivist* 58, no. 4 (1995): 454-456.
- . "Cultivating Archives: Meanings and Identities." *Archival Science* 12, no. 1 (2012): 19-33.
- . "Tacit Narratives: The Meaning of Archives." *Archival Science* 1, no. 2 (2001): 131-141.
- Khan, Kushal and Craig Locatis. "Searching through Cyberspace: The Effects of Link Display and Link Density on Information Retrieval from Hypertext on the World Wide Web." *Journal of the American Society for Information Science* 49, no. 2 (1998): 176-182.
- Kipp, Margaret E.I. "@toread and Cool: Subjective, Affective and Associative Factors in Tagging." In *Proceedings of the 36th Conference of the Canadian Association for Information Science*, 2008. http://www.caais-aci.ca/proceedings/2008/kipp_2008.pdf.

- Kipp, Margaret E.I. and D. Grant Campbell. "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices." In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*, 2006. <http://eprints.rclis.org/archive/00008315/>.
- Koltun, Lilly. "The Promise and Threat of Digital Options in an Archival Age." *Archivaria* 47, no. 1 (1999): 114–135.
- Koutrika, Georgia, Frans Adjie Effendi, Zoltan Gyöngyi, Paul Heymann, and Gector Garcia-Molina. "Combating Spam in Tagging Systems." In *Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web*, 2007. <http://dl.acm.org/citation.cfm?id=1244420>.
- Krause, Magia Ghetu and Elizabeth Yakel. "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid." *American Archivist* 70, no. 2 (2007): 282–314.
- Kunde, Nancy. "Getting It Done—Collaboration and Development of the Digital Records Conversion Standard." *American Archivist* 72, no. 1 (2009): 146–169.
- Lacher-Feldman, Jessica. "Making Friends and Fans: Using Facebook for Special Collections Outreach." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 54-64. Chicago: Society of American Archivists, 2011.
- Lazonder, Ard W., Harm J. A. Biemans, and Iwan G. J. H. Wopereis. "Differences Between Novice and Experienced Users in Searching Information on the World

Wide Web.” *Journal of the American Society for Information Science* 51, no. 6 (2000): 576–581.

Leadbeater, Charles and Debbie Powell. *We-Think: Mass Innovation, Not Mass Production*. London: Profile Books, 2008.

Lee, Chei Sian, Dion Hoe-Lian Goh, Khasfariyati Razikin, and Alton T.K. Chua.

“Tagging, Sharing and the Influence of Personal Experience.” *Journal of Digital Information* 10, no. 1 (2009).

<http://journals.tdl.org/jodi/index.php/jodi/article/view/275>.

Lee, Christopher A. and Helen Tibbo. “Where’s the Archivist in Digital Curation?

Exploring the Possibilities through a Matrix of Knowledge and Skills.”

Archivaria 72 (2011): 123-168.

Lee, Christopher A., ed. *I, Digital: Personal Collections in the Digital Era*. Chicago: Society of American Archivists, 2011.

Lewis, Michael. *Next: The Future Just Happened*. New York: W. W. Norton & Company, 2001.

Light, Michelle and Tom Hyry. “Colophons and Annotations: New Directions for the Finding Aid.” *American Archivist* 65 (2002): 24-41.

Lyons, Matthew. “K - 12 Instruction and Digital Access to Archival Materials.” *Journal of Archival Organization* 1, no. 1 (2002): 19–34.

Maher, William J. “The Importance of Financial Analysis of Archival Programs.” *Midwestern Archivist* 3, no. 2 (1978): 3–24.

March on Milwaukee Civil Rights History Project. University of Wisconsin-Milwaukee

Libraries. <http://www4.uwm.edu/libraries/digilib/march/index.cfm>.

Marchionini, Gary. "Information-Seeking in Full-Text End-User-Oriented Search

Systems: The Roles of Domain and Search Knowledge." *Library and Information*

Science Research 15, no. 1 (1993): 35–69.

Marchionini, Gary, Sandra Dwiggins, and Xia Lin. "Effects of Search and Subject

Expertise on Information-Seeking in a Hypertext Environment." In *Proceedings*

of the 53rd Annual Meeting of the American Society for Information Science, 129–

142. Washington, D.C.: American Society for Information Science, 1990.

Marchionini, Gary. "Information-Seeking Strategies of Novices Using a Full-Text

Electronic Encyclopedia." *Journal of the American Society for Information*

Science 40, no. 1 (1989): 54–66.

Marchionini, Gary. *Information Seeking in Electronic Environments*. Cambridge:

Cambridge University Press, 1995.

Marlow, Cameron, Mor Naaman, Danah Boyd, and Marc Davis. "HT06, Tagging Paper,

Taxonomy, Flickr, Academic Article, to Read." In *Proceedings of the Seventeenth*

Conference on Hypertext and Hypermedia, 2006.

<http://dl.acm.org/citation.cfm?id=1149949>.

Marshall, Catherine C. "Toward an Ecology of Hypertext Annotation." In *Proceedings of*

the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time

and Space, 40–49. New York: ACM, 1998.

- Martzoukou, Konstantina. "A Review of Web Information Seeking Research: Considerations of Method and Foci of Interest." *Information Research* 10, no. 2 (2005). <http://www.informationr.net/ir/10-2/paper215.html>.
- Mathes, Mathes. "Folksonomies: Cooperative Classification and Communication through Shared Metadata." 2004. <http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf>.
- Matusiak, Krystyna K. "Towards User-Centered Indexing in Digital Image Collections." *OCLC Systems & Services: International Digital Library Perspectives* 22, no. 4 (2006): 283–298.
- McCarthy, Paul H. "The Management of Archives: A Research Agenda." *American Archivist* 51, no. 1/2 (1988): 52–69.
- McClurken, Jeffrey W. "Waiting for Web 2.0: Archives and Teaching Undergraduates in a Digital Age." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 243–254. Chicago: Society of American Archivists, 2011.
- McCrea, Donna E. "Getting More for Less: Testing a New Processing Model at the University of Montana." *American Archivist* 69, no. 2 (2006): 284–290.
- McDonald, Sharon and Rosemary J. Stevenson. "Navigation in Hyperspace: An Evaluation of the Effects of Navigational Tools and Subject Matter Expertise on Browsing and Information Retrieval in Hypertext." *Interacting with Computers* 10, no. 2 (1998): 129–142.

- McKemmish, Sue and Michael Piggott, eds. *The Records Continuum: Ian Maclean and Australian Archives: First Fifty Years*. Clayton, Australia: Ancora Press, 1994.
- McKemmish, Sue. "Placing Records Continuum Theory and Practice." *Archival Science* 1, no. 4 (2001): 333–359.
- McKemmish, Sue, Shannon Faulkhead, and Lynette Russell. "Distrust in the Archive: Reconciling Records." *Archival Science* 11, no. 3–4 (2011): 211–239.
- Medina-Smith, Andrea. "Going Where the Users Are: The Jewish Women's Archive and Its Use of Twitter." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 65–74. Chicago: Society of American Archivists, 2011.
- Meissner, Dennis and Mark A. Greene. "More Application While Less Appreciation: The Adopters and Antagonists of MPLP." *Journal of Archival Organization* 8, no. 3–4 (2010): 174–226.
- Mendes, Luiz H., Jennie Quiñonez-Skinner, and Danielle Skaggs. "Subjecting the Catalog to Tagging." *Library Hi Tech* 27, no. 1 (2009): 30–41.
- Mercer Sabre, Jeannette and Susan Hamburger. "A Case for Item-Level Indexing: The Kenneth Burke Papers at The Pennsylvania State University." *Journal of Archival Organization* 6, no. 1–2 (2008): 24–46.
- Mitchell, William J. "Architectural Archives in the Digital Era." *American Archivist* 59, no. 2 (1996): 200–204.

Monks-Leeson, Emily. "Archives on the Internet: Representing Contexts and Provenance from Repository to Website." *American Archivist* 74, no. 1 (2011): 38-57.

Moore, Reagan W. "Building Preservation Environments with Data Grid Technology." *American Archivist* 69, no. 1 (2006): 139–158.

Morrison, P. Jason. "Tagging and Searching: Search Retrieval Effectiveness of Folksonomies on the World Wide Web." *Information Processing and Management* 44 (2008): 1562-1579.

Nesmith, Tom. "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice." *Archivaria* 60 (2005): 259-274.

———. "Seeing Archives: Postmodernism and the Changing Intellectual Place of Archives." *American Archivist* 65, no. 1 (2002): 24-41.

Nicholas, David, Paul Huntington, Hamid R. Jamali, and Tom Dobrowoiski. "Characterising and Evaluating Information Seeking Behaviour in a Digital Environment: Spotlight on the 'bouncer.'" *Information Processing & Management* 43, no. 4 (2007): 1085–1102.

Nimer, Cory and J. Gordon Daines. "What Do You Mean It Doesn't Make Sense? Redesigning Finding Aids from the User's Perspective." *Journal of Archival Organization* 6, no. 4 (2008): 216–232.

Nordlie, Ragnan "User revelation—A Comparison of Initial Queries and Ensuing Question Development in Online Searching and Human Reference Interaction." In *Proceedings of the 22nd Annual International ACM SIGIR Conference on*

Research and Development in Information Retrieval, 1999.

<http://dl.acm.org/citation.cfm?id=312624.312618>.

Nov, Oded, Mor Naaman, and Chen Ye. "Analysis of Participation in an Online Photo-Sharing Community: A Multidimensional Perspective." *Journal of the American Society for Information Science and Technology* 61, no. 3 (2010): 555–566.

O'Neill Adams, Margaret. "Analyzing Archives and Finding Facts: Use and Users of Digital Data Records." *Archival Science* 7, no. 1 (2007): 21–36.

Oliver, Gillian, Brenda Chawner, and Hai Ping Liu. "Implementing Digital Archives: Issues of Trust." *Archival Science* 11, no. 3–4 (2011): 311–327.

O'Sullivan, Catherine. "Diaries, On-line Diaries, and the Future Loss to Archives; Or, Blogs and the Blogging Bloggers Who Blog Them." *American Archivist* 68, no. 1 (2005): 53–73.

Ormond-Parker, Lyndon and Robyn Sloggett. "Local Archives and Community Collecting in the Digital Age." *Archival Science* 12, no. 2 (2012): 191–212.

Palmer, Joy "Archives 2.0: If We Build It, Will They Come?" *Ariadne* 60 (2009).
<http://www.ariadne.ac.uk/issue60/palmer>.

Palmer, Joy and Jane Stevenson. "Something Worth Sitting for? Some Implications of Web 2.0 for Outreach." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 1–21. Chicago: Society of American Archivists, 2011.

- Patel, Swapnesh C., Colin C. Drury, and Valerie L. Shalin. "Effectiveness of Expert Semantic Knowledge as a Navigational Aid within Hypertext." *Behaviour & Information Technology* 17, no. 6 (1998): 313–324.
- Pearce-Moses, Richard. "Janus in Cyberspace: Archives on the Threshold of the Digital Era." *American Archivist* 70, no. 1 (2007): 13–22.
- Peccatte, Patrick. "Liberating Archival Images: The PhotosNormandie Project on Flickr." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 148–158. Chicago: Society of American Archivists, 2011.
- Peters, Isabella. *Folksonomies: Indexing and Retrieval in Web 2.0*, translated by Paul Becker. Berlin: Walter de Gruyter, 2009.
- Piper, A.I. "Conducting Social Science Laboratory Experiments on the World Wide Web." *Library and Information Science Research* 20, no. 1 (1998): 5–21.
- Powell, Ronald R and Lynn Silipigni Connaway. *Basic Research Methods for Librarians*, 4th edition. Westport, Conn.: Libraries Unlimited, 2004.
- Prom, Christopher J. "User Interactions with Electronic Finding Aids in a Controlled Setting." *American Archivist* 67, no. 2 (2004): 234–268.
- Rafferty, Pauline and Rob Hilderley. "Flickr and Democratic Indexing: Dialogic Approaches to Indexing." *Aslib Proceedings* 59, no. 4/5 (2007): 397–410.

Ranger, Joshua. "More Bytes, Less Bite: Cutting Corners in Digitization." 2008.

<http://www.archivists.org/conference/sanfrancisco2008/docs/session701-ranger.pdf>.

Research Library Group. *Trusted Digital Repositories: Attributes and Responsibilities*.

Mountain View, CA: Research Libraries Group, 2002.

<http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.

Rogan, Mary Ellen. "The Wilson Project, an Archival Success Story." *Performance! The*

Newsletter of the Society of American Archivists' Performing Arts Roundtable,

Fall 2006 (2006).

<http://www.archivists.org/saagroups/performart/newsletter/PArtsNewsltr2006fal.pdf>.

Rorissa, Abebe. "A Comparative Study of Flickr Tags and Index Terms in a General

Image Collection." *Journal of the American Society for Information Science and*

Technology 61, no. 11 (2010): 2230–2242.

Ross, Steven M. and Gary R. Morrison. "Experimental Research Methods." In *Handbook*

of Research for Educational Communications and Technology, edited by D.H.

Jonassen. 1996. <http://aect.org/edtech/ed1/>.

Samouelian, Mary. "Embracing Web 2.0: Archives and the Newest Generation of Web

Applications." *American Archivist* 72, no. 1 (2009): 42–71.

Samuels, Helen W. "Improving Our Disposition: Documentation Strategy." *Archivaria*

33 (1991-1992): 125-140.

- . *Varsity Letters: Documenting Modern Colleges and Universities*. Chicago: Society of American Archivists, 1992.
- Sazedj, Peyman and H. Sofia Pinto. “Time to Evaluate: Targeting Annotation Tools.” In *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation*, 2005. <http://ceur-ws.org/Vol-185/semAnnot05-04.pdf>.
- Schaffer, Mirko Tobias. *Bastard Culture! User Participation and the Extension of Cultural Industries*. Amsterdam: Amsterdam University Press, 2011.
- Schindler, Amy. “A New Look for Old Information: Creating a Wiki to Share Campus History.” In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 191-202. Chicago: Society of American Archivists, 2011.
- Schmidt, Lisa. “Preserving the H-Net Email Lists: A Case Study in Trusted Digital Repository Assessment.” *American Archivist* 74, no. 1 (2011): 257–296.
- Schwartz, Joan M. and Terry Cook. “Archives, Records, and Power: The Making of Modern Memory.” *Archival Science* 2 (2002): 1-19.
- Sen, Shilad, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. “Tagging, communities, vocabulary, evolution.” In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, pp. 181-190. New York: ACM, 2006.
- Shepard, Elizabeth. “Digitizing a Photographic Collection in a Midsize Repository: A Case Study.” *Journal of Archival Organization* 2, no. 4 (2004): 67–82.

- Shepherd, Elizabeth, The Archival Education and Research Institute (AERI), and Pluralizing the Archival Curriculum Group (PACG). "Educating for the Archival Multiverse." *American Archivist* 74, no. 1 (2011): 69–101.
- Shilton, Katie and Ramesh Srinivasan. "Participatory Appraisal and Arrangement for Multicultural Archival Collections." *Archivaria* 63, no. 1 (2007): 87–101.
- Shirky, Clay. *Cognitive Surplus: Creativity and Generosity in a Connected Age*. New York: Penguin Press, 2010.
- Singer, Eleanor and Mick P. Couper. "Do Incentives Exert Undue Influence on Survey Participation? Experimental Evidence." *Journal of Empirical Research on Human Research Ethics* 3, no. 3 (2008): 49-56.
- Singer, Eleanor and Richard A. Kulka. "Paying Respondents for Survey Participation." In *Studies of Welfare Populations Data Collection and Research Issues*, edited by Michele Ver Ploeg, Robert A Moffitt, and Constance F Citro, 105-128. Washington, DC: National Academy Press, 2002.
- Sleeman, Patricia. "Notes and Communications It's Public Knowledge: The National Digital Archive of Datasets." *Archivaria* 58, no. 1 (2004): 173–200.
- Slotkin, Helen W. and Karen T. Lynch. "An Analysis of Processing Procedures: The Adaptable Approach." *American Archivist* 45, no. 2 (1982): 155–163.
- Šnuderl, Katja. "Tagging: Can User-Generated Content Improve Our Services?" *Statistical Journal of the IAOS* 25 (2008): 125–132.

- Speck, Jason G. "Protecting Public Trust: An Archival Wake-Up Call." *Journal of Archival Organization* 8, no. 1 (2010): 31–53.
- Speller, Edith. "Collaborative Tagging, Folksonomies, Distributed Classification or Ethnoclassification: A Literature Review." *Library Student Journal* 2 (2007).
<http://www.librarystudentjournal.org/index.php/ljsj/article/view/45>.
- Spiteri, Louise F. "The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog." *Information Technology and Libraries* 26, no. 3 (2013): 13–25.
- . "The Use of Folksonomies in Public Library Catalogues." *Serials Librarian* 51, no. 2 (2006): 75–89.
- Springer, Michelle, Beth Dulabahn, Phil Michel, Barbara Natanson, David Reser, David Woodward, and Helena Zinkham. *For the Common Good: The Library of Congress Flickr Pilot Project*. The Library of Congress, 2008.
http://www.loc.gov/rr/print/flickr_report_final.pdf.
- Steele, Tom. "The New Cooperative Cataloging." *Library Hi Tech* 27, no. 1 (2009): 68–77.
- Strom, Michael. "Texas-Sized Progress: Applying Minimum-Standards Processing Guidelines to the Jim Wright Papers." *Archival Issues* 29, no. 2 (2005): 105–112.
- Stvilia, Besiki and Corinne Jörgensen. "User-Generated Collection-Level Metadata in an Online Photo-Sharing System." *Library & Information Science Research* 31, no. 1 (2009): 54–65.

- . “Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr.” *Journal of the American Society for Information Science and Technology* 61, no. 12 (2010): 2477–2489.
- Sutcliffe, A. G., M. Ennis, and S. J. Watkinson. “Empirical Studies of End-User Information Searching.” *Journal of the American Society for Information Science* 51, no. 13 (2000): 1211–1231.
- Taormina, Mattie. “The Virtual Archives: Using Second Life to Facilitate Browsing and Archival Literacy.” In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 42-53. Chicago: Society of American Archivists, 2011.
- Theimer, Kate. “What Is the Meaning of Archives 2.0?” *American Archivist* 74, no. 1 (2011): 58–68.
- Theimer, Kate, ed. *A Different Kind of Web: New Connections between Archives and Our Users*. Chicago: Society of American Archivists, 2011.
- Townsend, Robert B. “Old Divisions, New Opportunities: Historians and Other Users Working with and in Archives.” In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 213-232. Chicago: Society of American Archivists, 2011.
- Trace, Ciaran B. and Andrew Dillon. “The Evolution of the Finding Aid in the United States: From Physical to Digital Document Genre.” *Archival Science* 12, no. 4 (2012): 501–519.

- Trant, Jennifer. "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept." *Art Museums: Proof of Concept. New Review of Hypermedia and Multimedia* 12, no. 1 (2006): 83–105.
- . "Tagging, Folksonomy and Art Museums: Early Experiments and Ongoing Research." *Journal of Digital Information* 10, no. 1 (2009).
<http://journals.tdl.org/jodi/index.php/jodi/article/view/270>.
- . *Tagging Folksonomy and Art Museums: Results of Steve.Museum's Research*, 2009. <http://www.archimuse.com/research/steve.html>.
- . "Studying Social Tagging and Folksonomy: A Review and Framework." *Journal of Digital Information* 10, no. 1 (2009).
<http://journals.tdl.org/jodi/index.php/jodi/article/view/269>.
- Triller, Malinda. "Double-Duty Blogging: A Reference Blog for Management and Outreach." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 203–212. Chicago: Society of American Archivists, 2011.
- Tyacke, Sarah. "Archives in a Wider World: The Culture and Politics of Archives." *Archivaria* 52, no. 1 (2001): 1–25.
- Upward, Frank, Sue McKemmish, and Barbara Reed. "Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures." *Archivaria* 72 (2011): 197–237.

- Uren, Victoria, Philipp Cimiano, José Iria, Siegfried Handschun, Maria Vargas-Vera, Enrico Motta, and Fabio Ciravegna. "Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art." *Web Semantics* 4, no. 1 (2006): 14–28.
- Van Ness, Carl. "Much Ado about Paper Clips: 'More Product, Less Process' and the Modern Manuscript Repository." *American Archivist* 73, no. 1 (2010): 129–145.
- Vardigan, Mary and Cole Whiteman. "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context." *Archival Science* 7, no. 1 (2007): 73–87.
- Vassileva, Julita. "A Task-Centered Approach for User Modeling in a Hypermedia Office Documentation System." *User Modeling and User-Adapted Interaction* 6, no. 2–3 (1996): 185–223.
- Vaughan, Jason. "Insights into the Commons on Flickr." *Libraries and the Academy* 10, no. 2 (2010): 185–214.
- Voigt, Tobi. "Is National History Day Ready for Web 2.0?" In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 233–242. Chicago: Society of American Archivists, 2011.
- von Ahn, Luis, Ruoran Liu, and Manuel Blum. "Peekaboom: A Game for Locating Objects in Images." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06*, 55–64. New York: ACM, 2006.

- Vuorikari, Riina. "Folksonomies, Social Bookmarking and Tagging: The State-of-the-Art." *Special Insight Reports*, 2007.
- http://insight.eun.org/shared/data/insight/documents/specialreports/Specia_Report_Folksonomies.pdf.
- Wang, Xianhua, Peter Liebscher, and Gary Marchionini. *Improving Information-Seeking Performance in Hypertext: Roles of Display Format and Search Strategy*. College Park: University of Maryland, 1988.
- Watson, Andrea, and P. Toby Graham. "CSS Alabama 'Digital Collection': A Special Collections Digitization Project." *American Archivist* 61, no. 1 (1998): 124–134.
- Weideman, Christine. "Accessioning as Processing." *American Archivist* 69, no. 2 (2006): 274–283.
- Weinberger, David. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York: Times Books, 2007.
- . *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York: Basic Books, 2011.
- Westbrook, Bradley D. "Prospecting Virtual Collections." *Journal of Archival Organization* 1, no. 1 (2002): 73–80.
- Westcott, Jezmyne, Alexandra Chappell, and Candace Lebel. "LibraryThing for Libraries at Claremont." *Library Hi Tech* 27, no. 1 (2009): 78–81.

- Wildemuth, Barbara M. "Descriptive Statistics." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 338-347. Westport, CT: Libraries Unlimited, 2009.
- . "Frequencies, Cross-Tabulation, and the Chi-Square Statistic." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 348-360. Westport, CT: Libraries Unlimited, 2009.
- . "Sampling for Extensive Studies." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 116-128. Westport, CT: Libraries Unlimited, 2009.
- Wildemuth, Barbara M. and Leo L. Cao. "Experimental Studies." In *Applications of Social Research Methods to Questions in Information and Library Science*, edited by Barbara M. Wildemuth, 105-115. Westport, CT: Libraries Unlimited, 2009.
- Withers, Charles W. J. and Andrew Grout. "Authority in Space?: Creating a Digital Web-Based Map Archive." *Archivaria* 61, no. 61 (2006): 27-46.
- Wojcik, Caryn. "Appraisal, Reappraisal, and Deaccessioning." *Archival Issues* 27, no. 1 (2002): 151-160.
- Xie, Iris and Edward Benoit III. "Search Result List Evaluation versus Document Evaluation: Similarities and Differences." *Journal of Documentation* 69, no. 1 (2013): 49-80.

- Xie, Iris. *Interactive Information Retrieval in Digital Environments*. Hershey, PA: IGI Pub., 2008.
- Xu, Zhichen, Yun Fu, Jianchang Mao, and Difu Su. "Towards the Semantic Web: Collaborative Tag Suggestions." *Collab. Web Tagging Workshop in Conjunction with the 15th International World Wide Web Conference*, 2006.
<http://semanticmetadata.net/hosted/taggingws-www2006-files/13.pdf>.
- Yakel, Elizabeth. "Balancing Archival Authority with Encouraging Authentic Voices to Engage with Records." In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 75-101. Chicago: Society of American Archivists, 2011.
- . "Inviting the User into the Virtual Archives." *OCLC Systems & Services* 22, no. 3 (2006): 159–163.
- Zach, Lisl and Marcia Frank Peri. "Practices for College and University Electronic Records Management (ERM) Programs: Then and Now." *American Archivist* 73, no. 1 (2010): 105–128.
- Zarro, Michael A. and Robert B. Allen. "User-Contributed Descriptive Metadata for Libraries and Cultural Institutions." In *Research and Advanced Technology for Digital Libraries*, edited by Mounia Lalmas et al., *Lecture Notes in Computer Science* 6273, 46-54. Berlin: Springer, 2010.
- Zhang, Jane. "Archival Representation in the Digital Age." *Journal of Archival Organization* 10, no. 1 (2012): 45–68.

———. “Original Order in Digital Archives.” *Archivaria* 74, no. 1 (November 16, 2012): 167–193.

———. “The Principle of Original Order & the Organization and Representation of Digital Archives.” Dissertation, Simmons College, 2010.

Zinkham, Helena and Michelle Springer. “Taking Photographs to the People: The Flickr Commons Project and the Library of Congress.” In *A Different Kind of Web: New Connections between Archives and Our Users*, edited by Kate Theimer, 102-115. Chicago: Society of American Archivists, 2011.

Zinn, Howard. “Secrecy, Archives, and the Public Interest.” *Midwestern Archivist* 2, no. 2 (1977): 14-27.

Zollers, Alla. “Emerging Motivations for Tagging: Expression, Performance, and Activism.” In *Proceedings of the 16th International World Wide Web Conference*, 2007. Retrieved from http://www2007.org/workshops/paper_55.pdf.

APPENDIX A: SAMPLE COLLECTION GROUPINGS

Type	Folder	Item	Link
Document Group 1: Hate Mail			
Doc	Groppi Papers, Box 8, Folders 3-6, Correspondence, Hate Mail	Anonymous letter, 1967 February 8	http://collections.lib.uwm.edu/u/?march,709
Doc	Groppi Papers, Box 8, Folders 3-6, Correspondence, Hate Mail	Anonymous letter, 1967 June 26	http://collections.lib.uwm.edu/u/?march,710
Doc	Groppi Papers, Box 8, Folders 3-6, Correspondence, Hate Mail	Anonymous letter, 1967 August 30	http://collections.lib.uwm.edu/u/?march,711
Doc	Groppi Papers, Box 8, Folders 3-6, Correspondence, Hate Mail	Anonymous letter, 1967 August 30	http://collections.lib.uwm.edu/u/?march,713
Doc	Groppi Papers, Box 8, Folders 3-6, Correspondence, Hate Mail	Anonymous letter, 1967 August 31	http://collections.lib.uwm.edu/u/?march,714
Document Group 2: Support Mail			
Doc	Groppi Papers, Boxes 1-4, Correspondence, Support Mail	Kenneth Croom letter, 1966 December 22	http://collections.lib.uwm.edu/u/?march,924
Doc	Groppi Papers, Boxes 1-4, Correspondence, Support Mail	Mike LaValle letter, 1967 July 25	http://collections.lib.uwm.edu/u/?march,928
Doc	Groppi Papers, Boxes 1-4, Correspondence, Support Mail	Waverly Davis letter, 1967 September 20	http://collections.lib.uwm.edu/u/?march,929
Doc	Groppi Papers, Boxes 1-4, Correspondence, Support Mail	Roger Tulin letter, 1967 September 20	http://collections.lib.uwm.edu/u/?march,930
Doc	Groppi Papers, Boxes 1-4, Correspondence, Support Mail	Leonard Mills letter, 1967 November 4	http://collections.lib.uwm.edu/u/?march,931
Document Group 3: Criticism Mail			

Doc	Groppi Papers, Boxes 5-7, Correspondence, Criticism Mail	Anonymous letter, 1967 September 17	http://collections.lib.uwm.edu/u?/march,1239
Doc	Groppi Papers, Boxes 5-7, Correspondence, Criticism Mail	Anonymous letter, 1967 August 30	http://collections.lib.uwm.edu/u?/march,1240
Doc	Groppi Papers, Boxes 5-7, Correspondence, Criticism Mail	Anonymous letter, 1967 August 5	http://collections.lib.uwm.edu/u?/march,1243
Doc	Groppi Papers, Boxes 5-7, Correspondence, Criticism Mail	Anonymous letter, 1967 August 13	http://collections.lib.uwm.edu/u?/march,1244
Doc	Groppi Papers, Boxes 5-7, Correspondence, Criticism Mail	Anonymous letter, 1967 August 15	http://collections.lib.uwm.edu/u?/march,1245
Photograph Group			
Photo	Groppi Papers, Photographs	Madison, Wisconsin, assembly chambers welfare protest, James Groppi center, 1969	http://collections.lib.uwm.edu/u?/march,650
Photo	Groppi Papers, Photographs	Madison, Wisconsin, demonstration at capitol protesting welfare cuts, 1969	http://collections.lib.uwm.edu/u?/march,651
Photo	Groppi Papers, Photographs	Madison, Wisconsin, James Groppi with raised fist, assembly chambers welfare protest, 1969	http://collections.lib.uwm.edu/u?/march,652
Photo	Groppi Papers, Photographs	Confrontation between Milwaukee police and the Milwaukee NAACP Youth Council, circa	http://collections.lib.uwm.edu/u?/march,653

1967-1968			
Photo	Gropi Papers, Photographs	NAACP march with James Gropi in the center, 1968	http://collections.lib.uwm.edu/u/?march,654
Photo	Gropi Papers, Photographs	James Gropi on witness stand, circa 1967-1968	http://collections.lib.uwm.edu/u/?march,655
Photo	Gropi Papers, Photographs	James Gropi, circa 1967-1968	http://collections.lib.uwm.edu/u/?march,656
Photo	Gropi Papers, Photographs	Madison, Wisconsin state capital welfare demonstration, 1969	http://collections.lib.uwm.edu/u/?march,657
Photo	Gropi Papers, Photographs	Meeting of NAACP commandos with James Gropi, circa 1967-1968	http://collections.lib.uwm.edu/u/?march,658
Photo	Gropi Papers, Photographs	Fair housing march, James Gropi center, 1967	http://collections.lib.uwm.edu/u/?march,659
Photo	Gropi Papers, Photographs	Saint Boniface public school boycott, James Gropi center, 1965	http://collections.lib.uwm.edu/u/?march,660
Photo	Gropi Papers, Photographs	James Gropi in back of police wagon, 1966	http://collections.lib.uwm.edu/u/?march,661
Photo	Gropi Papers, Photographs	The Freedom House on fire, 1967	http://collections.lib.uwm.edu/u/?march,662
Photo	Gropi Papers, Photographs	James Gropi and Vel Phillips on school bus, circa 1967-1968	http://collections.lib.uwm.edu/u/?march,663

Photo	Gropi Papers, Photographs	Stop bussing for segregation march, James Gropi center, 1968	http://collections.lib.uwm.edu/u?/march,664
--------------	------------------------------	--	---

APPENDIX B: PARTICIPANT KNOWLEDGE ASSESSMENT

The questions will be presented in a random order for each participant.

1. Vel Phillips was:
 - a) The first African-American woman elected to the Milwaukee Common Council
 - b) A professor at the University of Wisconsin-Milwaukee
 - c) The police chief of Milwaukee
 - d) A neighborhood watch leader
2. Which social club was picketed for its whites-only policy?
 - a) Wisconsin Club
 - b) Turner Hall
 - c) Eagles Club
 - d) Tripoli Shrine Temple
3. Which street was primarily used for the open housing march on August 28, 1967?
 - a) 6th Street
 - b) 16th Street
 - c) Capital Drive
 - d) North Avenue
4. What action did the Milwaukee mayor take following a racial disturbance in July 1967?
 - a) Shut down the inner city
 - b) Instructed police to arrest demonstrators
 - c) Announced a 24-hour city-wide curfew
 - d) No action
5. During 1967, the national media began referring to Milwaukee as:
 - a) Slum City
 - b) Selma of the North
 - c) Cream City
 - d) Deutsch-Athen
6. What building was burned out during the open housing marches?
 - a) Freedom House
 - b) Eagles Ballroom
 - c) St. Boniface Parish
 - d) Republican House
7. Who of the following was a major civil rights leader in Milwaukee?
 - a) Victor Berger
 - b) Harold Breier
 - c) Henry Maier

- d) James Groppi
8. In 1960, African-Americans accounted for what percentage of Milwaukee's population?
- a) 15%
 - b) 30%
 - c) 60%
 - d) 70%
9. Which of the following groups helped lead the civil rights movement in Milwaukee?
- a) NAACP Youth Council
 - b) Milwaukee Chamber of Commerce
 - c) Black Panthers
 - d) Republican National Committee
10. Who was mayor of Milwaukee during the civil rights movement?
- a) Daniel W. Hoan
 - b) John Bohn
 - c) Henry W. Maier
 - d) Frank P. Zeidler

APPENDIX C: PRE-QUESTIONNAIRE

Please complete the following information:

1. Age:

2. Gender:

Female
Male
Other

3. Race (select at least one):

White
Black
Hispanic or Latino
American Indian or Alaska Native
Asian/Indian subcontinent
Pacific Islander
Other

4. What is your religious affiliation, if any?

Protestant Christian
Roman Catholic
Evangelical Christian
Jewish
Muslim
Hindu
Buddhist
Atheist
Other:
Prefer not to state

5. What is the highest level of education you have completed?

Grammar school
High school or equivalent
Vocational/technical school (2 years)
Some college
Bachelor's degree
Master's degree
Doctoral degree
Professional degree (MD, JD, etc.)
Other

6. Are you currently a student?

Yes, full-time

Yes, part-time

No

7. Where do you currently reside?

[dropdown list of countries; if United States or Canada are selected then secondary dropdown list of state/province]

Please indicate your knowledge/experience level for the following categories by moving the available slider

8. *Knowledge and experience with computers*

Limited-----Very experienced

9. *Prior use of a digital collection*

Never-----Frequently

10. *Prior use of an archive*

Never-----Frequently

11. *Prior knowledge of social tagging*

Limited-----Very knowledgeable

12. *Prior use of social tagging*

Never-----Frequently

APPENDIX D: POST-QUESTIONNAIRE

Please answer the following questions regarding your tagging experience during the study

1. Please indicate your agreement/disagreement regarding the following statements based on your experiences with the study.

	strongly disagree	disagree	neither agree or disagree	agree	strongly agree
I found submitting tags easy					
If allowed, I would likely submit tags while using a digital archive in the future					
I enjoyed tagging documents					
I found tagging documents more difficult than tagging photographs					
I enjoyed tagging photographs					
I found tagging photographs more difficult than tagging documents					

2. What would make the tagging experience better? (open-ended)

3. Based on your experience, please indicate your agreement/disagreement regarding the following completions of the statement: When creating a tag, I considered...

	strongly disagree	disagree	neither agree or disagree	agree	strongly agree
How I would find the item					

How others would find the item

The content of the item

The item's format

The connection between items

**The accuracy of the provided
information**

The previous user's tags

My previous tags

4. What other considerations did you think of when creating your tags? Did these considerations change at all while you were tagging? (open-ended)

5. Please rate likelihood you would provide tags to a digital archive under each of the following conditions, from extremely unlikely to extremely likely.

	extremely unlikely	unlikely	neutral	likely	extremely likely
Archive requires you to create a user account and login to submit tags					
Archive offers recognition for tagging in newsletter or website					
Archive recognizes top taggers through social media (Facebook, Twitter, etc.)					
Archive provides non-monetary rewards for tagging (research assistance, archive tour, etc.)					
Archive allows you to anonymously submit tags					
Archive provides monetary rewards for tagging					

(photographic prints,
photocopies, discounted or free
membership, etc.)

6. Are there any other methods an archive could use to encourage tagging? (open-ended)

*Curriculum Vitae***Edward Benoit, III**

School of Information Studies
 University of Wisconsin-Milwaukee
 Milwaukee, WI 53201
 eabenoit@uwm.edu

Education

University of Wisconsin-Milwaukee, 2014

Ph.D. in Information Studies, Dissertation: "#MPLP: A comparison of domain novice and expert user generated tags in a minimally processed digital archives," supervised by Iris Xie.

University of Wisconsin-Milwaukee, 2009

Master of Library and Information Science

University of Wisconsin-Milwaukee, 2009

Master of Arts in History, Thesis: "A democracy of its own: Milwaukee's socialisms, difference and pragmatism," supervised by Aims McGuinness.

University of Wisconsin-Milwaukee, 2006

Bachelor of Arts in History, Magna Cum Laude, Honors in the Major, Capstone: "Georgette 'Dickey' Chapelle: A case study in photographs as historical sources."

Scholarly Activities

Publications

- Xie, I. & Benoit, E., III. (2013). Search result list evaluation versus document evaluation: Similarities and differences. *Journal of Documentation*, 69(1), 49-80.
- Benoit, E., III. (2011). Sub-field visualization: A multidimensional analysis of Web 2.0 authors. In E. Benoit, III (Ed.), *Proceedings of the 2011 Great Lakes Connections Conference*, 17-26. Retrieved from <http://digital.library.wisc.edu/1793/54429>.
- Benoit, E., III. (Ed.) (2011). *Proceedings of the 2011 Great Lakes Connections Conference*. Retrieved from <http://digital.library.wisc.edu/1793/54421>.
- Xie, I., Benoit, E., III., & Zhang, H. (2010). How do users evaluate individual documents? An analysis of dimensions of evaluation activities. *Information Research*, 15(4). Retrieved from <http://InformationR.net/ir/15-4/colis723.html>.

Conference Presentations

- Houston, B., Benoit, E., III, Dietz, B., Groth, J.E., & Noonan, D.W. (2014). The DAO of processing: Applying MPLP to electronic records workflows. Presented at the *Midwest Archives Conference*, Kansas City, MO.
- Benoit, E., III. (2014). Tagging MPLP: A comparison of novice and expert domain user generated tags in a minimally processed digital photographic archive. Presented at the *Visual Resource Association Conference*, Milwaukee, WI.

- Benoit, E., III, & Murillo, A. (2013). Building a collaborative archival research community. Presented at the *Archival Education and Research Institute*, Austin, TX.
- Benoit, E., III. (2012). Scrolls to scrolling: The shared heritage of digital collections. Presented at the *Sixth International Conference on the History of Records and Archives*, Austin, TX.
- Benoit, E., III., & Ramdeen, S. (2012). Wait, wait, don't tell me where I put that!: Research management software demonstration and discussion. Presented at the *Archival Education and Research Institute*, Los Angeles, CA.
- Benoit, E., III., Anderson, K.D., & Powell, A. (2012). Mind the gaps: Leveraging networks to learn. Presented at the *Midwest Archives Conference*. Grand Rapids, MI.
- Benoit, E., III. (2011). Social tagging on the Commons on Flickr: Comparing the Library of Congress with the remaining institutions. Presented at the *2011 School of Information Studies Student Research Symposium*, Milwaukee, WI.
- Benoit, E., III. (2011). Archival preservation and the Digital Millennium Copyright Act (DMCA): The need for relief action. Presented at the *Archival Education and Research Institute*, Boston, MA.
- Benoit, E., III. (2011). Sub-field visualization: A multidimensional analysis of web 2.0 authors. Presented at the *Great Lakes Connections Conference*, Milwaukee, WI.
- Benoit, E., III. (2011). Archival preservation and the Digital Millennium Copyright Act (DMCA): The need for relief action. Presented at the *Midwest Archives Conference*, St. Paul, MN.
- Benoit, E., III. (2010). Digital librarians' perceptions of social tagging, its potential use, benefits, and limitations. Presented at the *2010 School of Information Studies Student Research Symposium*, Milwaukee, WI.
- Xie, I., Benoit, E., III., & Zhang, H. (2010). How do users evaluate individual documents? An analysis of dimensions of evaluation activities. Presented at the *7th International Conference on Conceptions of Library and Information Science*, London, England.
- Benoit, E., III. (2010). Publishing with chains: A comparison of publishing agreements in LIS. Poster presented at the *7th International Conference on Conceptions of Library and Information Science*, London, England.
- Benoit, E., III. (2010). Tagging to access: Increasing access and reducing costs through integration of social tags and metadata in digital collections. Presented at the Doctoral Forum, *7th International Conference on Conceptions of Library and Information Science*, London, England.
- Benoit, E., III. (2008). Chicago's evolution of progress: Representations of the past, present and future at the 1893 Columbian Exposition and the 1933-34 Century of Progress Exposition. Presented at the *Phi Alpha Theta Biennial Convention*, Albuquerque, NM.
- Benoit, E., III. (2007). Chicago's evolution of progress: Representations of the past, present and future at the 1893 Columbian Exposition and the 1933-34 Century of Progress Exposition. Presented at the *Midwest Regional Conference*, Milwaukee, WI.

Benoit, E., III. (2005). Georgette 'Dickey' Chapelle: A case study in photography as history sources. Presented at the *Phi Alpha Theta Michigan Conference*, Marquette, MI.

Invited Speaker

Archival survival kit. Convening Great Lakes Culture Keepers Regional Institute.

Ziibiwing Center of Anishinabe Culture and Lifeways, Mt. Pleasant, MI, 2014.

Getting through graduate school. Curate Thyself: Defining and Cultivating an Academic Trajectory in Digital Curation. *University of North Carolina at Chapel Hill*, 2013.

Home movies in society and the historical record. Home Movie Day Preview Night. *University of Wisconsin-Milwaukee*, 2012.

Digital archiving and preservation. *University of Wisconsin-Milwaukee*, 2010, 2011, & 2012.

Expanding and contracting freedoms in American history. *University of Wisconsin-Milwaukee*, 2009.

Podcast creation, post-production, and publishing. *University of Wisconsin-Milwaukee*, 2009.

New acquisitions: Midnight Sun Baseball Classic and composite metal wood bat. Artifact spotlight at the *National Baseball Hall of Fame*, Cooperstown, NY, 2007.

Eight Men Out: The fictionalized version of the 1919 Black Sox scandal versus historical fact. Hollywood versus History Film Series, Milwaukee, WI, 2007.

Forest Home Cemetery: A walking tour of Milwaukee history. Phi Alpha Theta, Milwaukee, WI, 2005.

Works in Progress

Xie, I., & Benoit, E., III. History of the University of Wisconsin-Milwaukee's School of Information Studies Ph.D. program. Book chapter in progress.

Benoit, E., III. Archival preservation and the Digital Millennium Copyright Act (DMCA): The need for relief action. Journal article in progress.

Benoit, E., III. Social tagging on the Commons on Flickr: Comparing the Library of Congress with the remaining institutions. Journal article in progress.

Benoit, E., III. Possibilities and perceptions: A study of digital librarians' and social tagging. Journal article in progress.

Benoit, E., III. E pluribus unum: Defining digital archives. Journal article in progress.

Seidel, E. *Sketches from My Life: The Autobiography of Emil Seidel* E. Benoit, III (Ed.). Manuscript in progress.

Research Project Experience

School of Information Studies (2013-2014). University of Wisconsin-Milwaukee Graduate School Review of the School of Information Studies' Ph.D. Program (compiled and analyzed data, and created report narrative).

Xie, I. (2012). Digital Library Teaching Resource (DLTRe). [Http://dltre.sois.uwm.edu](http://dltre.sois.uwm.edu) (designed and constructed Drupal-based website).

Mahoney, M.S. (2011). *Histories of Computing*. T. Haigh, (Ed.). Cambridge, MA: Harvard University Press (compiled, formatted, and edited works of Mahoney; created diagrams and tables; and assisted with final editing).

- Smiraglia, R. & van den Heuvel, C. (2010-2011). "Idea Collider research project"
(compiled literature review and participated in theoretical model development).
- Walker, T. (2009) "A comparative study of seventeenth-century theatrical dance music at the Viennese court," ongoing research project (constructed aggregate database for analysis).

Teaching Experience

Adjunct Instructor

School of Information Studies, *University of Wisconsin-Milwaukee*, 2012-present
Arrangement & Description in Archives (Online, Fall 2013, Spring 2014)
Digital Libraries (Onsite, Fall 2013, Spring 2013, Spring 2012)
Information Resources for Research (Onsite, Fall 2012)
Fieldwork in Archives and Manuscripts (Online, Spring 2014)

Teaching Assistant

School of Information Studies, *University of Wisconsin-Milwaukee*, 2010-2011
Courses: Introduction to Information Science; Information Access and Retrieval;
and Digital Libraries.

Department of History, *University of Wisconsin-Milwaukee*, 2006-09

Courses: American Cultures: Africans, Europeans, and Indian Nations; American History: 1877 to the Present; The 1960's in the United States: A Cultural History; and World History since 1500.

Awards

School of Information Studies Teaching Reward, *School of Information Studies, University of Wisconsin-Milwaukee*, Fall 2012, Spring 2013, Fall 2013.

Doctoral Research Award Grant Opportunity, *School of Information Studies, University of Wisconsin-Milwaukee*, 2012, 2013, 2014.

Archival Education and Research Institute Scholarship, 2011, 2012, 2013, 2014.

Conference Presentation Travel Grant, *School of Information Studies, University of Wisconsin-Milwaukee*, 2010, 2011, 2012, 2013, 2014.

Chancellor's Scholarship, *University of Wisconsin-Milwaukee*, 2010, 2011, 2012, 2013.

Teaching Assistantship/Project Assistantship, *School of Information Studies, University of Wisconsin-Milwaukee*, 2009-2013.

Great Lakes National STEM Scholarship, 2012.

Graduate School Travel Grant, *University of Wisconsin-Milwaukee*, 2010, 2012.

Beta Phi Mu, *University of Wisconsin-Milwaukee*, 2012.

Midwest Archives Conference Student Scholarship, 2011.

Dean's Scholarship, *School of Information Studies, University of Wisconsin-Milwaukee*, 2010, 2011.

Chancellor's Fellowship, *University of Wisconsin-Milwaukee*, 2009, 2010.

Theodore Saloutos Graduate Research Fellowship, *University of Wisconsin-Milwaukee*, 2007.

Graduate Research Grant, *University of Wisconsin-Milwaukee*, 2007.

Graduate Paper Honorable Mention, *Phi Alpha Theta Midwest Regional Conference*, 2007.

Frank and Peggy Steele Internship, *National Baseball Hall of Fame*, 2007.

Teaching Assistantship, Department of History, *University of Wisconsin-Milwaukee*, 2006-09.

Phi Alpha Theta, *University of Wisconsin-Milwaukee*, 2005.

Phi Kappa Phi, *University of Wisconsin-Milwaukee*, 2005.

Academic and Professional Service

Archival Studies Program Assistant, *School of Information Studies, University of Wisconsin-Milwaukee*, 2013-present.

Student Day Planning Committee, *Archival Education and Research Institute*, 2013-present.

Social Studies of Information Research Group, Executive Committee, *School of Information Studies, University of Wisconsin-Milwaukee*, 2012-present.

Social Studies of Information, Website Manager, 2012-present.

Special Website Management Volunteer, *Special Interest Group in Computers, Information and Society*, 2010-present.

SOIS Doctoral Student Organization, Executive Officer, *University of Wisconsin-Milwaukee*, 2012-2013.

Doctoral Committee, Ph.D. Rep., *School of Information Studies, University of Wisconsin-Milwaukee*, 2009-2011.

SOIS Doctoral Student Organization, President, *University of Wisconsin-Milwaukee*, 2010-2011.

Faculty Council, Ph.D. Rep, *School of Information Studies, University of Wisconsin-Milwaukee*, 2009-2011.

Editor, SOIS PhD Newsletter, *School of Information Studies, University of Wisconsin-Milwaukee*, 2012.

Chief Editor, SOIS PhD Newsletter, *School of Information Studies, University of Wisconsin-Milwaukee*, 2011.

Connections 2011 Conference Committee, Co-chair, 2010-2011.

Conference Blogger, Midwest Archives Conference, 2011.

Faculty Search and Screen Committee, Ph.D. Rep., *School of Information Studies, University of Wisconsin-Milwaukee*, 2011.

Assistant to the Editor, Book Reviews, *Digest of Middle East Studies*, 2010.

SOIS Graduate Student Organization, Executive Committee, *University of Wisconsin-Milwaukee*, 2009-2010.

Pedagogical and Professional Issues in Library and Information Science Lecture Series, Organizer, *University of Wisconsin-Milwaukee*, 2009-10.

Bargaining Committee, *Milwaukee Graduate Assistants Association*, 2007-2009.

Election Committee, *Milwaukee Graduate Assistants Association*, 2008.

Midwest Regional Conference, Organizer, *University of Wisconsin-Milwaukee*, 2007.

Phi Alpha Theta, Delta Phi, President, 2005-2008.

Professional Experience

Milwaukee Art Museum, Milwaukee, WI, Curatorial Intern, Audio/Visual Archive, 2009-10.

National Baseball Hall of Fame, Cooperstown, NY, Recorded Media Archive Intern, 2007.

Waukesha County Historical Society and Museum, Waukesha, WI, Archives Assistant, 2006.

Professional Affiliations

American Society for Information Science & Technology, since 2008.

Midwest Archives Conference, since 2011.

Society of American Archivists, since 2009.