

#MPLP Part 1: Comparing Domain Expert and Novice Social Tags in a Minimally Processed Digital Archives

Edward Benoit III

ABSTRACT

This study explores the potential for controlling/mediating the supplemental meta-data from user-generated tags through inclusion of domain-expert user-generated tags. The study was a mixed-methods, quasi-experimental design using a sample collection of fifteen documents and fifteen photographs. Sixty participants divided based on assessed prior domain knowledge tagged the sample collection. An open-coding analysis found six major categories and six subcategories. T-tests and chi-square tests found statistically insignificant or very weak associations between tags and domain knowledge. The study recommends inclusion of both expert and novice tags with each group demonstrating different qualities and serving different purposes.

© Edward Benoit III.



KEY WORDS

Social tagging, Participatory archives, MPLP

The emergence of Web 2.0 in the past decade provided a dynamic, interactive space where users collaborate, customize their information space, and engage with traditional information providers. As part of the Web 2.0 transition, social tagging within digital collections gained increased interest.¹ 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.² Users distrust the tags generated by other general users; however, they would consider using information created by so-called expert researchers and users.³

The previous decade also saw the introduction and implementation of Greene and Meissner's "More Product, Less Process" (MPLP) and minimal processing.⁴ Applied to digital archives, the minimal processing technique prioritizes the collection as a whole over individual items, specifically regarding metadata. The online collections provide only minimal metadata, or what Casey Davis and Sadie Roosa called "minimum viable metadata," typically at the series or folder level.⁵ 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 online users demand more description and access points.⁶

Reaching out to the same users for assistance and asking 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.⁷ This study proposes categorizing the taggers rather than the tags; specifically, permitting users who are subject-area experts (hereafter referred to as expert users) to tag the collections. It theorizes that expert users provide more reliable tags, meeting the needs of institutions and improving access to the collections.

This is the first of two articles presenting and discussing the results of a mixed-methods, quasi-experimental research project focused on tag generation within a sample minimally processed digital archives. The first article explores the potential for controlling/mediating the supplemental metadata from user-generated tags through inclusion of domain-expert user-generated tags by addressing the following research question and hypotheses:

RQ: What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archives?

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

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

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

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

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

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

The subsequent article compares the tags generated within a minimally processed collection with the existing item-level metadata from the sample collection. Additionally, the second article explores how both the tags and the existing metadata correspond with existing users' search terms.⁸

Literature Review

Understanding the placement of this article 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. The exploration of social tagging began broadly with research on Web-based tagging, mainly for personal use.⁹ The research shifted to include tagging within traditional information retrieval systems such as databases,¹⁰ online public access catalogs (OPACs),¹¹ and digital libraries.¹² Rather than focusing on the systems, many studies examined the tags and taggers themselves. This literature discussed an equally wide variety of topics as above, including taggers and their motivations for tagging,¹³ how familiarity with tagging affects the quality of tags,¹⁴ the wide range of categories of tags,¹⁵ their internal organization,¹⁶ and how tags develop.¹⁷

Perhaps the most promising tagging applications focus on digital collections, with many of these studies conducted by practitioners rather than researchers. The small number 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.¹⁸ 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,¹⁹ proposed methodological metrics,²⁰ highlighted case studies,²¹ explored the experiences of The Commons' participating institutions,²² and compared the tags of the Library of Congress with other Flickr-based institutions.²³

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 it increases access points. For example, Scott R. Anderson and Robert B. Allen viewed 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."²⁴ The reliability and authority of the metadata decrease, however, since the metadata is no longer strictly controlled.

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 showed the quantitative information and did not engage the users' experience or linguistically analyze the tags produced through coding.²⁵ Kevin Andreano highlighted the potential of social tagging within film archives that can be difficult to access since many archival collections remain poorly described.²⁶ Robert Townsend recognized the importance of tagging and other Web 2.0 applications for building and/or strengthening the archivist/user relationship.²⁷ Townsend also suggested opening collections to tagging and that increasing the number of digital archives available would provide evidence for future budget and funding meetings.

Social tagging is not without problems. Several researchers discuss the entropic nature of tags and tagging systems, such as variability within spellings and punctuation, and compound tag creation.²⁸ Social tags can also replicate information already provided. In an initial analysis of YouTube tags, Wooseob Jeong found that a high rate (46%) of tags was already included in the titles.²⁹ 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.³⁰

As such, digital librarians and archivists remain reluctant to allow tags and other user-generated content within their collections.³¹ While they are concerned about possible tag irregularities (i.e., misspellings, compound tag construction, etc.), profanity or spam issues are most troubling, although occurrences of profanity within tagging on sites such as Flickr are extremely rare.³² Georgia Koutrika et al. highlighted two related trends within tagging spam, specifically the creation of malicious tags intended to misdirect either a user or the system and so-called promotional tagging, where a content creator applies unrelated but popular tags to an item to increase viewing.³³

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 or inconsistencies within the tags. Marieke Guy and Emma Tonkin recommended posting best practices or a tutorial for users to view along with a combination of manual and automatic cleaning of existing tags.³⁴ Others suggested 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.³⁵ Finally, Zhichen Xu et al. commended a combination of approaches, including real-time algorithms, which highlight statistical outlier tags for possible deletion, tag weighting, and manually moderating tags.³⁶

The social tagging research, as a whole, appears well developed through its exploration of tagging with information retrieval (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 that address these concerns. Additionally, tagging in digital archives has not received as much research attention as tagging in digital libraries because of the lack of major tagging projects related to archives. This article addresses the gaps in both the archival and tagging literature by examining the use of expert-user-generated tags and by testing a possible quality-control mechanism for the tags requiring limited oversight by the archivist.

Definitions

The relative infancy and dynamic nature of born-digital and digitized records precludes a clear, concise, and universally agreed-upon definition of digital archives. 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. This article 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 this article, therefore, a *digital archives* is defined and limited to curated online collections of digitized materials selected from a single or multiple existing physical archival collection(s) that adhere to the archival principles of provenance and original order, and are, at a minimum, arranged and described following contemporary best archival practices. This definition

excludes collections of born-digital 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 discussed research project fulfilled the specified characteristics as it contained digitized correspondence and photographs selected as representative of an existing physical collection, and it maintained the physical collection’s arrangement and description through aggregation into compound digital objects (similar to folder-level arrangement).

Methodology

A mixed-methods, quasi-experimental design best addressed the research question and hypotheses by focusing on tag generation for a sample minimally processed digital archive. Table 1 provides an overview of the data-collection methods and analysis of the research question and each hypothesis.

Table 1. Research Question, Hypotheses, Associated Data Collection, and Analysis

Research Question/Hypotheses	Data Collected	Data Analysis
RQ: What are the similarities and differences between tags generated by expert and novice users in a minimally processed digital archive?	Prequestionnaire	Descriptive statistics
	Tags generated by expert and novice users (at least one tag for 30 items per participant)	Open-coding, descriptive statistics
H1: 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
H2: The number of photographic tags generated in a minimally processed digital archive is affected by a user’s domain knowledge.		Independent-samples t-tests
H3: The number of document tags generated in a minimally processed digital archive is affected by a user’s domain knowledge.		Independent-samples t-tests
H4: 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
H5: 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
H6: The proportion of document 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

SAMPLE COLLECTION

This study used selections from an existing digital collection to create a sample digital archives for the experiment. The creation of a sample collection derived from an existing collection created a comfortable setting and interface for participants during the data collection, thereby strengthening the internal validity of the data. Rather than randomly sampling from a single collection, the sample collection used a critical case-sampling technique. A random sample would not necessarily include items previously used within the existing digital collection and would therefore limit the amount of existing metadata for comparison with the tag 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.”³⁷ The selection procedure prioritized the format over content and included a combination of handwritten documents, typed documents, and photographic images.

The sample collection included 30 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 13 archival collections with a wide range of formats including audio, documents (handwritten and typed), photographs, and moving images. Additionally, the collection includes both personal and organizational records. 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.”³⁸

The personal papers of one of the main leaders of the Milwaukee movement, Father James Groppi, are included within March on Milwaukee and were selected as the sole source for the sample collection’s records because they contain materials in multiple formats. 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. Each of the four series/subseries of records was uploaded into a CONTENTdm-hosted digital collection with each grouping only displaying a set of shared minimal metadata (see Table 2).

SAMPLE POPULATION AND PARTICIPANT DEMOGRAPHICS

The data were generated by 60 participants divided equally through purposive sampling based on domain knowledge of the civil rights movement in Milwaukee. The overall population group focused on the metropolitan Milwaukee area because, in the real world, users from the region would most

Table 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			

likely access March on Milwaukee. Participants were limited to those over 18 years old; however, no additional exclusion criteria were enforced.

Participants were recruited through various methods including online postings, fliers, and directed invitations. 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. To increase the response rate, and because participation in the study required a time commitment of about 1.5 to 2 hours, each participant was compensated \$15 upon completion of the study.

Interested participants completed a prequestionnaire and were assigned to the novice or expert group unless the designated group reached its quota of 30 participants. The knowledge level or expertise of a given participant was determined through completion of a brief 10-question multiple-choice on the civil rights movement in Milwaukee. The author researched and developed the assessment questions based on prior knowledge of the topic and the subject matters of the sample collection materials. Additionally, an independent researcher knowledgeable on the subject reviewed the assessment tool, which was also tested by several colleagues with a variety of knowledge levels.

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. By dismissing intermediate users, the study avoided drawing conclusions from minuscule differences between those scoring a 4 and a 5. Among the 60 participants, the expert group had a mean score of 7.57 ($n = 30$) with the novice group providing a mean of 2.77 ($n = 30$).

Participants provided demographic information and self-assessed their computer literacy level, experience with digital collections and archives, and social tagging using a visual analog scale (VAS). A VAS presents a continuum from one extreme to the other without subdivision markings. For example, a VAS measuring self-assessment of experience with digital collections would include “no experience” at one end of a line and “very experienced” at the other with a movable slider. Participants then indicate their level by moving the slider from “no experience” to “very experienced.” The resulting data are reported on a scale of 0–100. According to Ben Hasson and Bengt B. 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.³⁹ Likert scales offer respondents a limited set of answers, such as strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.

The 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 participants were female, with a similar gender balance for both expert and novice groupings. Most participants came from either Wisconsin or Illinois (48.3%), although 21 states and the District of Columbia were represented in the study.

The majority of participants racially identified only as white (60%), while four participants (6.7%) indicated both white and nonwhite racial identifiers since participants could select multiple racial groupings. Excluding participants who identified as partially white, 33.3% of all participants were from nonwhite racial groupings. When compared with 2012 U.S. Census racial estimates for Wisconsin and Illinois combined, the participants closely reflected the real-world racial composition of the states.⁴⁰ The 2012 estimates provide a 69.1% to 30.9% racial division between white and nonwhite groupings, whereas the participants comprised a 66.7% to 33.3% racial division.

Table 3 reports the medians and means of the VAS scores for experts, novices, and the combination of both groups. Individual Mann-Whitney U tests were run to determine any differences in participants' self-assessed areas (prior use of digital collections, archives, and social tagging; knowledge of social tagging; and computer experience) between experts and novices. For all five areas,

Table 3. Average VAS Scores from Prequestionnaire

	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		

the distribution of levels for experts and novices was not similar, as assessed by visual inspection. All tests indicated a lack of statistically significant difference based on domain knowledge groupings, thus the participants represented a fairly homogeneous sample limiting any influence of these variables on the resulting data.

DATA COLLECTION METHODS AND PROCEDURES

Participant data collection during the study occurred in three phases: participant prequestionnaire, tag generation, and participant postquestionnaire. This article mainly focuses on data generated during the second phase. Following prequestionnaire 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.

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 subcollections

(documents and photographs) individually with the ability to move between records within the subcollection.

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 subcollections 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 participants to virtually “approve” or “thumbs up” previous submissions. The required instructional video also directed participants to provide only 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, they spent an estimated 1 to 1.5 minutes per item for a total of 1 to 1.5 hours for the tagging activity.

DATA ANALYSIS

Overall, the data analysis combined several approaches in qualitative and quantitative methods, thereby alleviating the limitations of one method with the strengths of another. A portion of the data analysis relied on multiple statistical analyses, therefore requiring clear delineations of the variables investigated. The independent variable for all statistical analyses was 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 comprised the independent variable. Membership in each of the domain knowledge groups was based on participants’ scoring during the prequestionnaire assessment; however, the knowledge level (and independent variable) was considered nominal since the assessment scores were used only to determine group membership and not to differentiate knowledge levels between members of the same group.

The qualitative tag analysis relied on grouping the tags into categories and subcategories. Although coding schemes exist from previous studies, this

Table 4. Coding Scheme Categories and 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

study developed a new coding scheme based on an open coding of the data. The application of open coding allowed “the categories and names for categories to flow from the data,” rather than forcing the data into structured silos.⁴¹ While the open coding method allowed the data to speak for themselves, the resulting analysis cannot be easily compared with previous studies.

Since the coding process required 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 4 lists and provides a definition for each category and subcategory. Further discussion of the categories occurs in the results section below.

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, 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 those of the researcher, resulting in a strong intercoder reliability of 0.954 based on Ole Holsti's reliability formula of $\frac{2M}{N_1 + N_2}$.⁴² 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, a very high level of agreement existed 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.⁴³

Part of the research question tested the association between the independent variable and the number of tags generated (dependent variable) in total, for the photograph set alone, 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 H_1-H_3 . A second portion of the research question explored a possible association between the independent variable and type or category of tag created (dependent). In this instance, the dependent variable was nominal, requiring chi-square tests for association (H_4-H_6).

Results

The following section discusses the results of the study related to the scope of the research question 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.

NUMBER OF TAGS GENERATED BY EXPERT AND NOVICE PARTICIPANTS

Combined, the participants generated a wide range of tags, from the required minimum of 30 to 1,031 tags created by one participant. 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 5 presents the aggregate tag counts by format and users including the number of unique tags. Figures 1 and 2 chart the number of tags generated by each participant divided by format.

Table 5. 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 recalculations, Experts, *n* = 28, Novices, *n* = 29.

At first glance, novice users appeared 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. Removing these outliers reduced 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). Testing for distribution normality determines the appropriate statistical analysis to use in each case. 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 ± 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 any differences in the three tag categories (all tags, photographic tags, and document tags) between experts and novices using *H*₁-*H*₃. Homogeneity of variances existed for experts

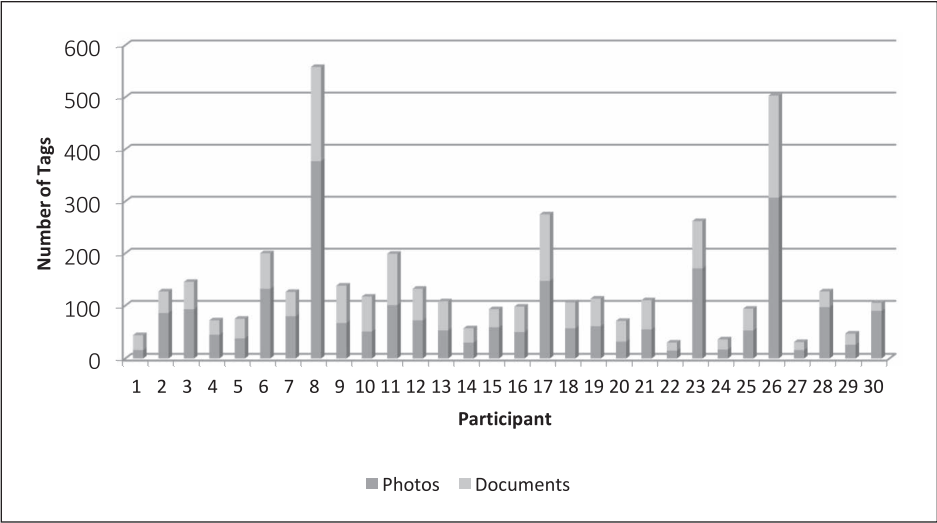


FIGURE 1. Expert Tag Counts by Format

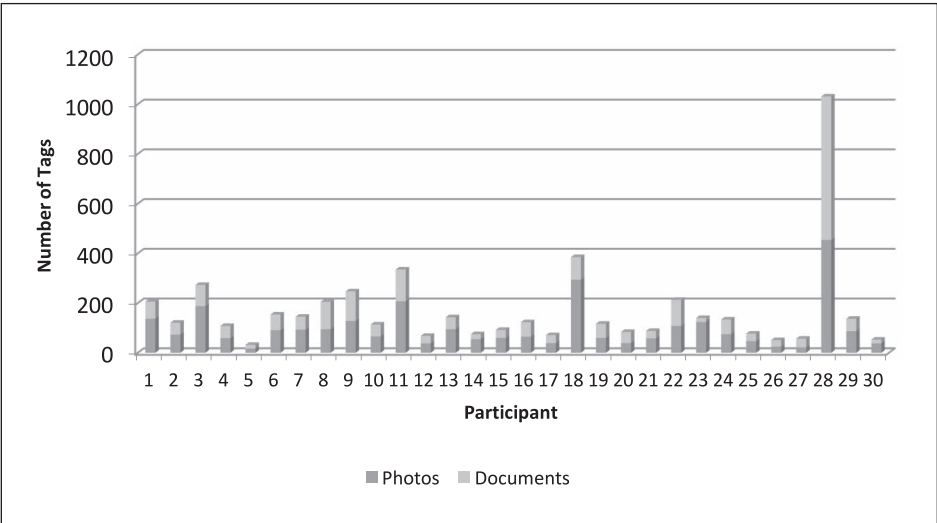


FIGURE 2. Novice Tag Counts by Format

and novices, as assessed by Levene’s test for equality of variances, for all tags ($p = .165$), photographic tags ($p = .185$), and document tags ($p = .376$). No statistically significant difference existed 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 no 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 no 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 generated. Both groups averaged above the minimum of 30 tags, indicating that most participants did not merely consider the minimum requirements for the study. Additionally, both experts and novices produced more tags for the documents than for 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 did the novices for both photographs and documents.

TAG CATEGORIES AND TYPES

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 intercoder reliability testing phase. The final coding scheme, therefore, included 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 Table 4 for definitions of the categories with examples).

The first major category, replication of metadata, included tags that duplicated information already presented to the user in the minimal metadata for each item. The minimal metadata included information from the following fields: Title, 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.⁴⁴

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 a difference existed 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 formats of the items themselves. 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.”

The majority of tags across all items served as subjects in some fashion (49.49%), thereby creating the largest major category of tags. The subject tags category contained two subcategories: general and specific. Tags in the former subcategory identified objects, places, or people with common nouns, such as “police,” “demonstrators,” or “youth.” The latter tags used proper nouns and provided more specific information, such as “Milwaukee Police,” “CORE,” or “NAACP Youth Council.” Additionally, the subject-specific tags included dates for the photographs and documents.

The combined tag analysis found 25.64% as subject-general and 23.85% as subject-specific. Although the combination of photograph and document tags split evenly between general and specific subjects, separating the formats revealed an intriguing difference. The photograph tags’ general/specific gap was 13.1 percentage points in favor of general (25.24%/12.14%), whereas the document tags’ general/specific gap was 6.22 percentage points in favor of specific (25.93%/32.15%). The formats themselves explain the difference as the documents provided participants directly with proper nouns to use as tags within the letters through simple transcription, while the photographs required more prior knowledge or interpretation for specific identification.

Tags placed into the content-summary category were those that described and/or summarized what was going on in the photograph or document. These tags comprised 16.32% of all tags, 16.35% of photograph tags, and 8.53% of document tags. Similar to the subject tags, the nature of the formats revealed the format disparity. Since the photographs required more interpretation, they produced a higher percentage 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 and represented 13% of all tags. Often these tags focused on the civil rights movement or a theme within the movement, such as “race,” “segregation,” “nonviolence,” “solidarity,” or “religion.” Although these terms appear as tags within other categories, their use in relation to the specific item tagged 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, these tags are identification-general. Participants used the same tag for Photograph 11, and because “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.

The penultimate major category included tags containing an emotional response to one of the objects. 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. The original coding scheme did not include the last category; however, after discussion with the outside coder used for intercoder 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. Participant E26 provided 14 incorrect tags (51.9%) and Participant N23 added 9 incorrect tags (33.3%); combined, the two participants accounted 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 3 different tags incorrectly (“catholic hate,” “criticism,” and “hate mail”) all within the support mail letters.

Table 6 provides the categorical distribution for photograph, document, and all tags; Figure 3 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 differed 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 aggregates Subject-Specific (32.15%), Subject-General (25.93%), and Replication of Metadata (20.95%). The close relationship between the aggregate and document-specific categorizations was primarily caused by the higher number of document tags (compared to photograph tags) influencing the aggregate level.

Table 6. 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%

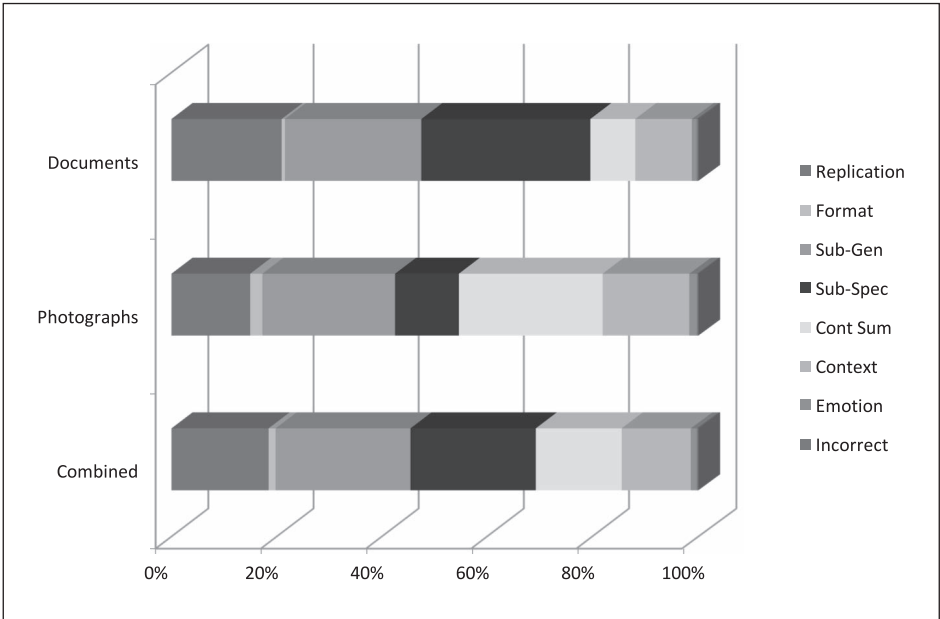


FIGURE 3. Comparison of Expert and Novice Tag Category Percentages by Format

SIMILARITIES AND DIFFERENCES BETWEEN EXPERT AND NOVICE 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

Table 7. Number and Percentage 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%

and documents revealed some initial similarities and differences (see Table 7 and Figures 4 and 5). The main similarities with both expert and novice tags highlight 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 section. Novices provided twice the number 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.

A chi-square test for association was conducted between domain group (expert/novice) and tag category to test the significance of expert and novice tag differences for all items based on H_4 , 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. A statistically significant association existed between domain group and tag category, $\chi^2(7) = 77.149$, $p < .0005$.⁴⁵ The association, however, was very weak, Cramer’s $V = 0.091$.

Dividing the tags by format was necessary to best explore the similarities and differences between expert and novice tags. Photographs and documents elicited different responses from experts and novices (see Table 8 and Figures 6 and 7). Novices’ photographic tags focused more on general subject terms, while experts provided more content-summary and context tags for photographs by taking a broader 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

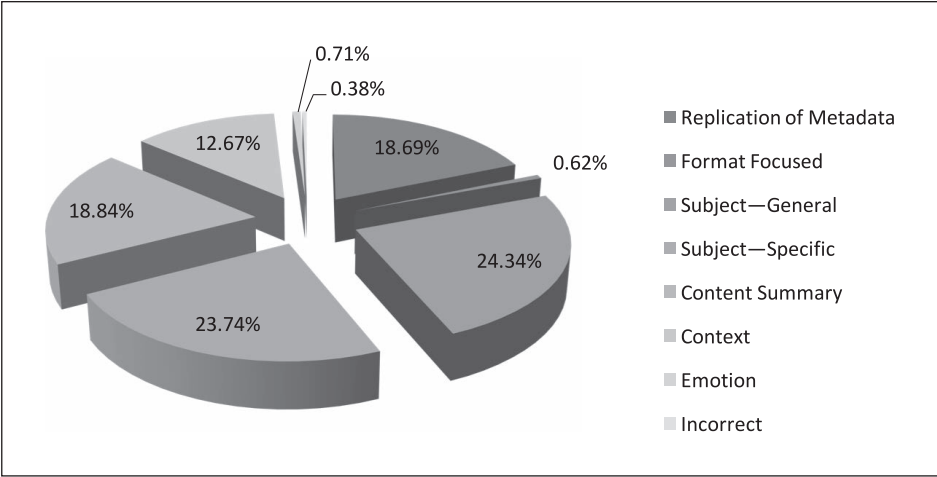


FIGURE 4. All Expert Tags by Category

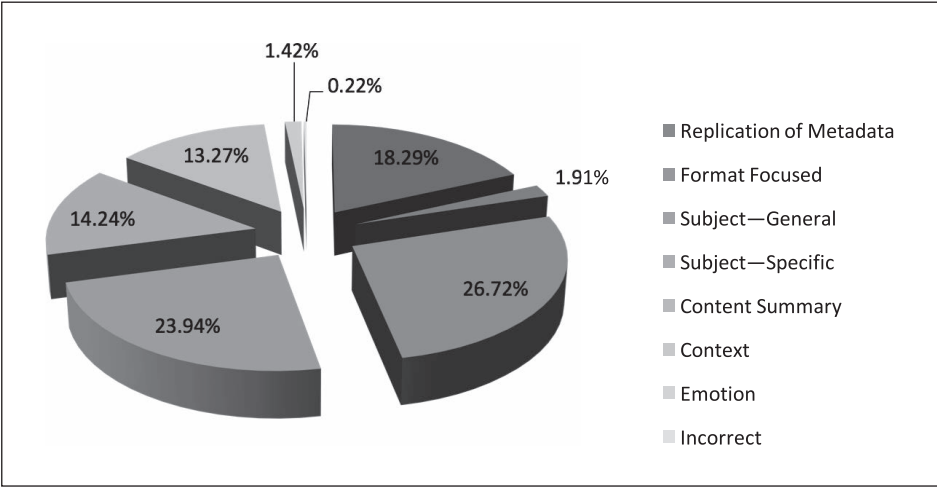


FIGURE 5. All Novice Tags by Category

approaches toward the photographs. Novices, having little domain knowledge background, attempted to identify individual parts of a photograph: a crowd, a library, a banner, a baton. Experts, on the other hand, identified what was going on in the captured scene: dissent, demonstration for racial justice, black-white solidarity.

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

Table 8. Number and Percentage of Expert and Novice Photography 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%

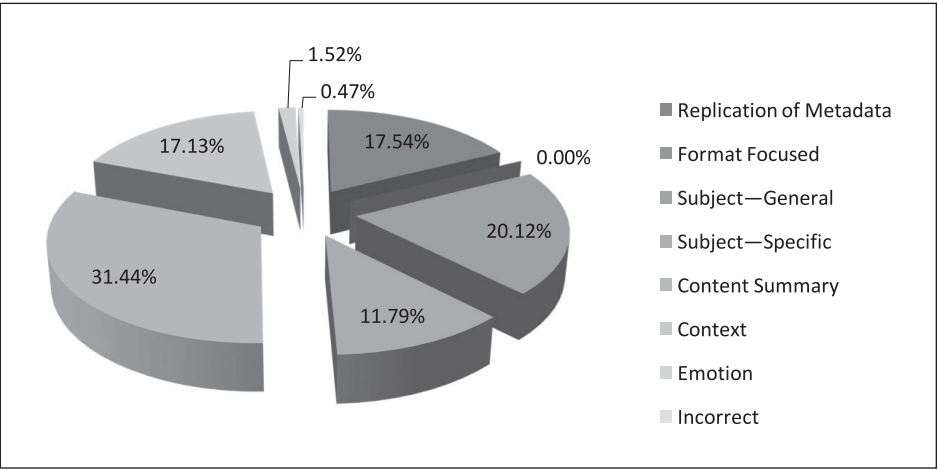


FIGURE 6. Expert Photograph Tags by Category

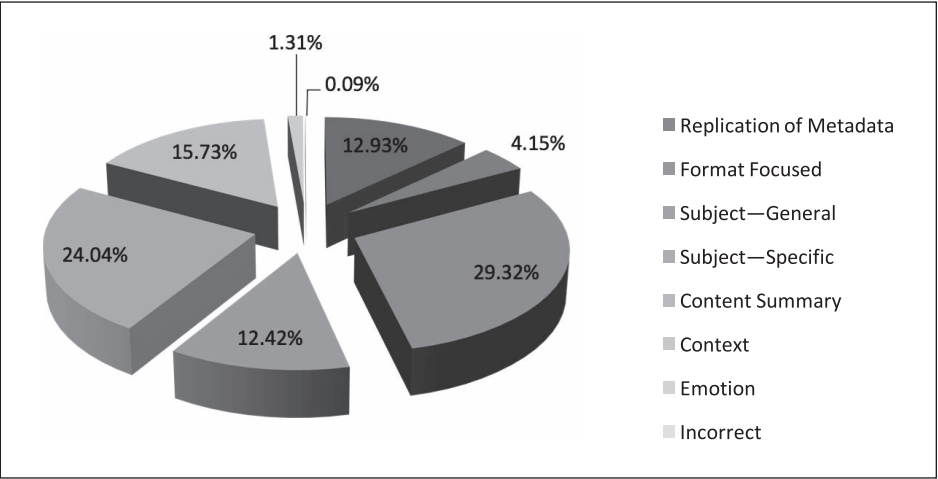


FIGURE 7. Novice Photograph Tags by Category

Table 9. Number and Percentage 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%

A chi-square test for association was conducted between domain group (expert/novice) and tag category to test the significance of expert and novice tag differences for photographs based on 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. As it was the only expected count below five, the chi-squared analysis could still be run. A statistically significant association existed between domain group and tag category, $\chi^2(7) = 142.043$, $p < .0005$.⁴⁶ The association, however, was weak, Cramer’s $V = 0.192$ (although stronger than the analysis of all tags).

The document tags offer a slightly different picture than the photographic tags (see Table 9 and Figures 8 and 9). In general, novices found the documents easier than photographs when it came to locating specific subjects, as 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 compared to 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 did experts for documents.

When compared within their own domain groupings, the experts created more unique tags (685) than did the novices (579). A cross-group comparison of unique tags found 295 terms in both groups’ unique tag lists. The experts

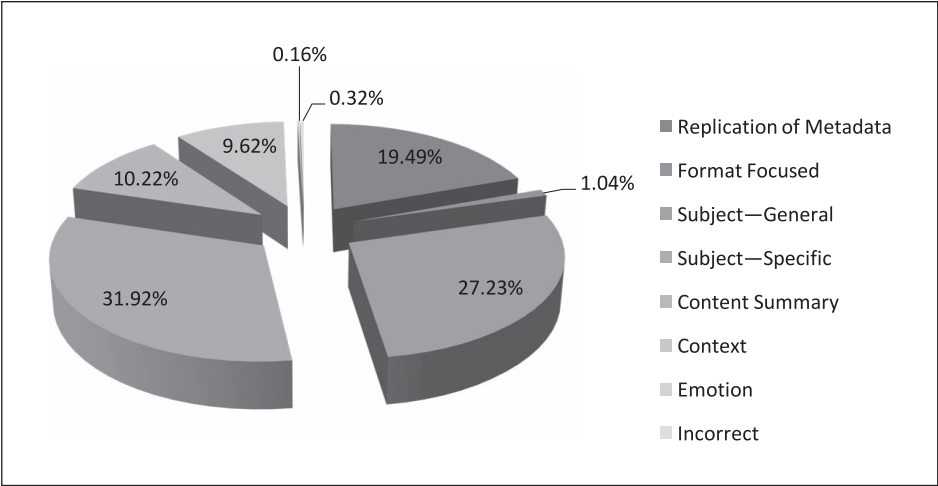


FIGURE 8. Expert Document Tags by Category

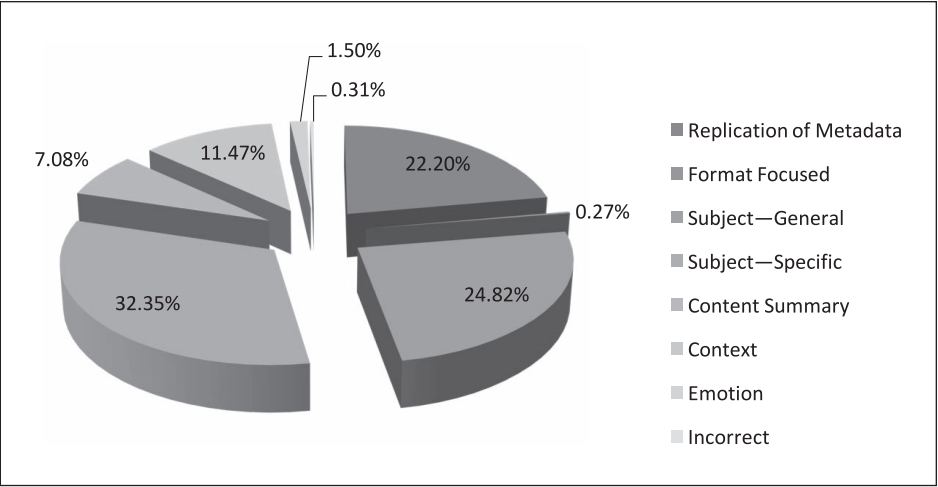


FIGURE 9. Novice Document Tags by Category

created 404 unique tags that 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 to test the significance of expert and novice tag differences for documents based on 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. A statistically significant association existed between domain group and tag category, $\chi^2(7) = 67.889$, $p < .0005$.⁴⁷ The association, however, was weak, Cramer’s $V = 0.112$ (stronger than the analysis of all tags, but weaker than the photograph tags).

All tested hypotheses for the research question indicated a statistically significant association between domain group and coded tag category. The associations were all relatively weak based on low Cramer'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 increasing the number of tags) could see the categorical differentials increase and strengthen the statistical association.

Discussion and Conclusion

The findings for the research question and the tested hypotheses indicate minute differences between expert and novice participants' tags with either statistically insignificant or very weak associations with domain knowledge groupings. The data shed light on several areas. The 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 focused on descriptions of the potential benefits of user participation or engagement rather than empirical testing. Studies by Andrew Flinn, Alexandra Eveleigh, or Isto Huvila, for example, encouraged the expansion of archival engagement through public collaboration throughout the archival processes.⁴⁸ Although these previous studies occasionally used 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. This study's findings offer 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 diverse interpretations of archival materials suggested by participatory archival research. Furthermore, the findings also reinforce Max Evans's discussion of relieving archives of the temporal and fiscal burdens of increased collections by "acting as partners" or "organizing agents" with users for item-level descriptions.⁴⁹

The findings also answer calls for additional research into the content created by users, and specifically how it could be integrated or used to supplement archival description.⁵⁰ The low number of incorrect tags within the study's findings also reinforces Joy Palmer's argument to treat users as "peer collaborators . . . rather than outside interlopers."⁵¹

Some of the study's results do not reflect previous work. For example, it did not find as many personal or emotional tags as previous tagging studies have, perhaps indicating participants considered others' use of the tagged object rather than their own personal use.⁵² 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 spamlike tagging behavior noted by Koutrika et al.⁵³ 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 Elizabeth Yakel's case study of social tagging of the Hague City Archives.⁵⁴ Additionally, the level and breadth of the description offered by the generated tags meets users' needs and desires as described by Jodi Allison-Bunnell, Elizabeth Yakel, and Janet Hauck's previous research on helpful metadata elements and users' opinions of Web 2.0 tools within digital archives.⁵⁵ The study addresses the users' reliability concerns through the lack of incorrect tags. The study also addresses Joyce Celeste Chapman's concerns regarding "the ability of the average Internet user to leave un-moderated content."⁵⁶ Although the data indicate concerns are not necessary, the emphasis must be changing the users' perception of tags through outreach and increasing the number of tags they see within digital archives.

The largest implications of the study's findings relate specifically to the application of tags within a minimally processed digital archives. In the introduction, the researcher proposes using prior domain knowledge as an indicator of tagging quality and, 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 does not suggest 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 people, places, objects, and time periods within the photographs and documents.

The lack of large variations between experts and novices indicates negative results for the study. 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 as 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 that 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.

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, as 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 its minimally processed digital collections, it 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's inclusion of both as major categories of tags. A major tagging concern from previous studies was the potential (or likelihood) of incorrect tags. The researcher addressed this concern by including incorrect tags within its coding analysis and found it to be the least occurring category throughout formats and domain groups, with only 27 occurrences of incorrect tags out of 9,278 tags (0.29%). Similar to the replication of the metadata problem, the lack of incorrect tags reaffirms previous findings, but at slightly lower rates.⁵⁷ The influence of tagging conditions, specifically the limited number of taggers and nonnatural 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.⁵⁸ 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.

Finally, the 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 social tagging with confidence toward its benefits rather than with unwarranted hesitation or fearfulness.

LIMITATIONS AND FUTURE DIRECTIONS

The results, implications, and limitations of this research project 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, taggers' ages, and the division of researchers and nonresearchers. Similarly, future studies should include additional archival formats to better compare tagging efficiency and efficacy. Formats such as audio and moving images may produce different results as they would require increased attention from the participants (due to the nature of the formats themselves).

The project used a nonnatural 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 current project, the longitudinal version's results would be more directly applicable for real-world digital archives.

Although the article'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 by 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 materials description and access points over time, and by not limiting the tagging to specific users, archives will continue striving for inclusiveness of opinions and perspectives rather than return to the exclusionary past.

NOTES

¹ 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: ACM, 2006), 151–54, 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): 283–98; Michelle Springer et al., *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; Jennifer Trant, "Exploring the Potential for Social Tagging and Folksonomy 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), 102–15.

² Scott R. Anderson and Robert B. Allen, "Envisioning the Archival Commons," *The American Archivist* 72, no. 2 (2009): 383–400; Adam Crymble, "An Analysis of Twitter and Facebook Use by the Archival Community," *Archivaria* 70 (2010): 125–51; Magia Ghetu Krause and Elizabeth Yakel,

- "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *The 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): 159–63.
- ³ 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.
- ⁴ Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68, no. 2 (2005): 208–63.
- ⁵ Sadie Roosa, "Understanding What Users Need to Understand Us (and Our Data)" (presentation, Annual Meeting of the Association of Moving Image Archivists, Portland, Oregon, November 18–21, 2015); and Casey Davis, "Navigating Copyright to Provide Access and Use" (presentation, Annual Meeting of the Association of Moving Image Archivists, Portland, Oregon, November 18–21, 2015).
- ⁶ Kate Cruikshank, Caroline Daniels, Dennis Meissner, Naomi L. Nelson, and Mark Shelstad, "How Do We Show You What We've Got? Access to Archival Collections in the Digital Age," *Journal of the Association for History and Computing* 8, no. 2 (2005), <http://hdl.handle.net/2027/spo.3310410.0008.203>; and Torou Elena, Akrivi Katifori, Costas Vassilakis, George Lepouras, and Constantin Halatsis, "Historical Research in Archives: User Methodology and Supporting Tools," *International Journal on Digital Libraries* 11, no. 1 (2010): 25–36.
- ⁷ Edward Benoit III and Amanda L. Munson, "Proceed with Caution: Changing Practitioner Perceptions of Social Tagging within Digital Collection, 2010–2016," submitted to *First Monday*.
- ⁸ Edward Benoit III, "#MPLP Part 2: Replacing Item-level Metadata with User-Generated Social Tags," *The American Archivist* 81, no. 1 (2018), forthcoming.
- ⁹ 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 2006), 726, 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* 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. 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: ACM, 1998), 40–49, 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): 1562–79; 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), <https://pdfs.semanticscholar.org/68f6/7e520e2e9e1622ecb86de701d8ecdaf74988.pdf>; Victoria Uren et al., "Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art," *Web Semantics* 4, no. 1 (2006): 14–28.
- ¹⁰ 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: ACM, 2007), 971–80, 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.
- ¹¹ 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 and Alton Y. K. Chua, "A Study of Web 2.0 Applications in Library Websites," *Library & Information Science Research* 32, no. 3 (2010): 203–11; Jonathan Furner, "User Tagging of Library Resources: Toward a Framework for System Evaluation" (presentation, World Library and Information Congress: 73rd IFLA

- General Conference and Council, Durban, South Africa, 2007), https://www.researchgate.net/profile/Jonathan_Furner/publication/238393988_User_tagging_of_library_resources_Toward_a_framework_for_system_evaluation/links/00b4953276369ba83a000000.pdf; Dimitris Gavrili, 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, 2008), 148–59, 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.
- ¹² 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), <http://www.archimuse.com/research/steve.html>; Jennifer 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.
- ¹³ 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, [doi.acm.org/10.1145/1124772.1124782](https://doi.org/10.1145/1124772.1124782); Alla Zollers, "Emerging Motivations for Tagging: Expression, Performance, and Activism," in *Proceedings of the 16th International World Wide Web Conference* (2007).
- ¹⁴ 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>.
- ¹⁵ 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>.
- ¹⁶ Pauline Rafferty and Rob Hilderley, "Flickr and Democratic Indexing: Dialogic Approaches to Indexing," *Aslib Proceedings* 59, nos. 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.
- ¹⁷ 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://rdcu.be/vcWF/>; 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://eprints.rclis.org/11748/1/kipp_2008.pdf.
- ¹⁸ 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"; 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."
- ¹⁹ 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 (2010): 2477–89; 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): 2230–42; 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): 555–66.
- ²⁰ 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.
- ²¹ 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.
- ²² Vaughan, “Insights into the Commons on Flickr.”
- ²³ Benoit III, “Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions.”
- ²⁴ Benoit III, “Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions,” 400.
- ²⁵ Edmunson-Morton, “Talking and Tagging: Using CONTENTdm and Flickr in the Oregon State University Archives.”
- ²⁶ 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.
- ²⁷ Robert B. Townsend, “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*, ed. Kate Theimer (Chicago: Society of American Archivists, 2011), 213–32.
- ²⁸ 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.”
- ²⁹ 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>.
- ³⁰ Wooseob Jeong, “Is Tagging Effective?—Overlapping Ratios with Other Metadata Fields,” *International Conference on Dublin Core and Metadata Applications* (2009): 31–39.
- ³¹ Benoit III, “Digital Librarians’ Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations.”
- ³² Benoit III, “Digital Librarians’ Perceptions of Social Tagging, Its Potential Use, Benefits, and Limitations”; Benoit III, “Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions.”
- ³³ Georgia Koutrika et al., “Combating Spam in Tagging Systems,” *Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web*, 2007, 57–64, <http://dl.acm.org/citation.cfm?id=1244420>.
- ³⁴ Guy and Tonkin, “Folksonomies.”
- ³⁵ Vuorikari, *Folksonomies, Social Bookmarking and Tagging: The State-of-the-Art*.
- ³⁶ Zhichen Xu et al., “Towards the Semantic Web: Collaborative Tag Suggestions,” *Collab. Web Tagging Workshop in Conjunction with the 15th International World Wide Web Conference*, 2006, <http://www.ambuehler.ethz.ch/CDstore/www2006/www.rawsugar.com/www2006/13.pdf>.
- ³⁷ Gary T. Henry, *Practical Sampling* (Newbury Park, Calif.: Sage, 1990), 21.
- ³⁸ March on Milwaukee Civil Rights History Project, <http://collections.lib.uwm.edu/cdm/landingpage/collection/march>.
- ³⁹ 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): 178–92. See also Robert F. DeVellis, *Scale Development: Theory and Applications* (Newbury Park, Calif.: Sage, 1991).
- ⁴⁰ United States Census Bureau, “State and County QuickFacts,” <https://web.archive.org/web/20140719104633/http://quickfacts.census.gov/qfd/states/00000.html>.
- ⁴¹ Hsiu-Fang Hsieh and Sarah E. Shannon, “Three Approaches to Qualitative Content Analysis,” *Qualitative Health Research* 15, no. 9 (2005): 1279.
- ⁴² Ole R. Holsti, *Content Analysis for the Social Sciences and Humanities* (Reading, Mass.: Addison-Wesley, 1969).

- ⁴³ 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, Conn.: Libraries Unlimited, 2009), 338–47; 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, Conn.: Libraries Unlimited, 2009), 348–60.
- ⁴⁴ 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.
- ⁴⁵ The p-value is 1.5455×10^{-14} .
- ⁴⁶ The p-value is 1.8965×10^{-27} .
- ⁴⁷ The p-value is 3.941×10^{-12} .
- ⁴⁸ 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>; Alexandra Eveleigh, "Welcoming the World: An Exploration of Participatory Archives" (paper presented at the International Conference on Archives, Brisbane, Australia, 2012), <http://www.gosbook.ru/system/files/documents/2012/11/13/ica12Final00128.pdf>; and Isto Huvila, "Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management," *Archival Science* 8, no. 1 (2008), 15–36.
- ⁴⁹ Max J. Evans, "Archives of the People, by the People, for the People," *The American Archivist* 70, no. 2 (2007), 397.
- ⁵⁰ Joy Palmer, "Archives 2.0: If We Build It, Will They Come?," *Ariadne* 60 (2009), <http://www.ariadne.ac.uk/issue60/palmer>.
- ⁵¹ Palmer, "Archives 2.0."
- ⁵² Kipp, "@toread and Cool: Subjective, Affective and Associative Factors in Tagging"; and Golder and Huberman, "Usage Patterns of Collaborative Tagging Systems."
- ⁵³ Koutrika et al., "Combating Spam in Tagging Systems."
- ⁵⁴ Yakel, "Inviting the User into the Virtual Archives."
- ⁵⁵ Allison-Bunnell, Yakel, and Hauck, "Researchers at Work."
- ⁵⁶ Joyce Celeste Chapman, "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids," *Journal of Archival Organization* 8, no. 1 (2010), 24.
- ⁵⁷ Benoit III, "Social Tagging on the Commons on Flickr: Comparing the Library of Congress with the Remaining Institutions."
- ⁵⁸ Jeong, "Does Tagging Really Work?" and Jeong, "Is Tagging Effective?—Overlapping Ratios with Other Metadata Fields."

ABOUT THE AUTHOR



Edward Benoit III is assistant professor at the School of Library and Information Science at Louisiana State University. He is the coordinator of both the archival studies and cultural heritage resource management MLIS specializations. He received an MA in history, an MLIS, and a PhD in information studies from the University of Wisconsin–Milwaukee. His research focuses on participatory and community archives, nontraditional archival materials, and archival education. He is the founder and director of the Virtual Footlocker Project that examines the personal archiving habits of twenty-first-century soldiers in an effort to develop new digital capture and preservation technologies to support their needs.