

Thomas S. Richey. Context Entities of Flickr Tags: Classes of Information Captured by Social Tagging. A Master's paper for the M.S. in L.S. degree. April, 2011. 44 pages.
Advisor: Christopher Lee

This study explores the content entities, or classes of information, captured by tags assigned to images posted to Flickr's The Commons site. Seven institutions participating in The Commons were examined: The New York Public Library, The Washington, D.C. Public Library, The University of Washington, Michigan State University, Cornell University, The Field Museum, and the George Eastman House. Tables that collected tags from these institutions and their subsequently assigned context entities revealed the kind of information expressed by users of Flickr's social tagging site.

Through a review of the literature and by examining data from Web traffic sites, it is shown that tagging is a popular form of describing images and that professionals across information fields generally accept it as a supplement to access points created by controlled vocabularies. From an analysis of Flickr tags, it is shown that objects and places are the most common contextual entities captured by user-generated tags. By implication, it is suggested what gaps in contextual entities professionals should fill for users, and vice-versa. This paper concludes with suggested improvements for the study and areas where social tagging will continue to engage and empower users of image sites and other services compatible with tagging functionality.

Headings:

Web 2.0 – Evaluation

Metadata – Folksonomy

Photograph Collections

CONTEXT ENTITIES OF FLICKR TAGS:
CLASSES OF INFORMATION CAPTURED BY SOCIAL TAGGING

by
Thomas S. Richey

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

Chapel Hill, North Carolina

April, 2011

Approved by:

Christopher Lee

TABLE OF CONTENTS

Acknowledgments.....	2
Introduction.....	3-5
Literature Review.....	6-22
Methodology	23-29
Findings.....	30-33
Limitations and Suggestions for Further Research	34-36
Conclusion	37-38
Works Cited	39-42

Acknowledgments

Many thanks to my advisor, Christopher (Cal) Lee, and my family for their patience and direction.

Introduction

This study aims to analyze how users of digital image collections describe images through tagging. Specifically, this study proposes to answer the following questions: What type of descriptive information do users of social media sites for images tend to provide? Or, what do the tags that users create say about the object described? Creating tags is an open process, not predefined by a descriptor set, such as the elements used in Dublin Core. This means that tagging can offer flexibility but also create ambiguity, inconsistency, and sometimes plain nonsense. Critics have responded to the advent of simple hypertext tagging with mixed feelings. The opening of description to users of digital image collections has been alternately described by some critics as creating a “cult of the amateur”¹ and by other commentators as encouraging user participation that is empowering, democratic, and knowledge-enhancing. One vocal critic, Andrew Keen, laments the rise of “the “noble amateur” and his or her threat to reliable information and professionalism. However, a major grant-funded study, *Steve: The Museum Social Tagging Project*, found that most museum professionals express optimism for and acceptance of the adoption of tagging functionality in digital museum collections. Moreover, a rising number of postmodern critics of archives have commented how traditional classification systems and access points restrict alternative ways of thinking about describing information. For example, a popular proponent of crowdsourcing, Clay

¹Keen, A. (2007). *The cult of the amateur: How today's internet is killing our culture*. New York, NY: Doubleday/Currency.

Shirky, argues that “the ‘exclusive’ nature of existing controlled vocabularies [impedes] their overall usability,” suggesting that “current schemes are incapable of reflecting the transient nature of knowledge and therefore the demands of the modern information user.”² Neither controlled vocabularies nor tags are perfect; as Krystyna Matusiak remarks that “the gap between user language and controlled vocabularies applied in indexing have been identified as a major problem in providing intellectual access to images.”³ However, social tagging has infiltrated the archive and will likely remain there. It therefore seems advantageous to understand what kind of information these tags impart.

Despite criticisms and inherent problems (synonyms, homonyms, homographs, tags with only personal relevance, and a generally high potential for “noise” in search results, among other concerns) with social tagging, its adoption for online digital collections allows users to create descriptive terms potentially overlooked by static and hierarchical descriptive schemas. Raya Fidel’s article “User-centered Indexing” calls for more research in users’ search behaviors so that information scientists can develop an automated indexing system that responds to user demands: “The static nature of intellectual indexing requires users to interact with retrieval systems to improve the results of their searches . . . therefore, indexing processes that are dynamic are more promising as user-centered indexing methods because they may tailor indexing to the requirements of each request” (575). Although in this context Fidel refers to computer-automated indexing, her suggestion to improve intellectual indexing through adopting

²Macgregor, G., McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5), 291-300.

³Matusiak, K.K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services: International Digital Library Perspectives*. 22(4), 283-298.

“dynamic” processes is relevant to user tagging, a characteristically adaptive way of describing digital objects. This study seeks to understand what a “collective intelligence” can offer when it creates a body of entities constituting a dynamic indexing system.

Literature Review

Tagging: Overview and Functions

Before examining the kind of information that user-generated tags transmit, it will be useful to attempt to define what tagging means and to describe what it does. To do so, one must place tagging in the context of Web 2.0 technological development. Despite its ubiquity in popular and academic literature, the use of the term Web 2.0 sometimes imparts a nebulous meaning. In an attempt to clarify its significance, Tim O'Reilly identifies common sites and services to distinguish "Web 2.0" from "Web 1.0," as sketched in Figure 1.

Web 1.0		Web 2.0
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

Figure 1. Attributes of Web 1.0 compared to Web 2.0.

For people already familiar with the given sites and services of Web 1.0 versus those of Web 2.0, the differences will be readily apparent. However, for those with limited

experience, O'Reilly summarizes the essential difference by stating that the new Web seizes the potential to “harness collective intelligence.”⁴ This “hive mind” characteristic of Web 2.0 differs from Web 1.0 in that users previously navigated a World Wide Web that mostly constituted a read-only experience, without any or much functionality for posting personal or perceived global meaning. This capacity for user input opens the Web to multiple interpretations and distances it from previously unchallenged assertions.

To place this altered Web in the context of postmodern discourse on “textuality,” Web 2.0 enables the creation of multiple narratives that collectively question grand, monolithic narratives and other such assertions of historical truth. Terry Cook has written extensively on the need for transparency, metadocumentation, and user-empowering functionality in archives,⁵⁶ as well as Michelle Light and Tom Hyry.⁷ By opening a dialogue on documents and information to all interested parties, Web 2.0 helps achieve such goals. David Weinberger (2007) remarks that the previous limits on organizing and sharing information “not only limited our vision, [but] they have also given the people who control the organization of information more power than those who create the information.”⁸ Web 2.0 and its potential to share control of information is therefore commonly heralded as the democratization of the World Wide Web, an assertion charged

⁴O'Reilly, T. (2005, September 30). What is web 2.0: Design patterns and business models for the next generation of software. Retrieved from <http://owl.english.purdue.edu/owl/resource/560/10/>

⁵Cook, T. (2001). Archival science and postmodernism: New formulations for old concepts. *Archival Science*, 1(1), 3-24.

⁶Cook, T. & Schwartz, J. (2002). Archives, records, and power: From (postmodern) theory to (archival) performance. *Archival Science*. 2(3-4), 171-185.

⁷Hyry, T. & Light, M. (2002). Colophons and annotations. *American Archivist*. 65(2), 216-230.

⁸Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York, NY: Times Books.

with controversy but in part accurate when one considers the increased number of people now involved in the creation of information on the Web.

Tagging, a specific Web 2.0 tool, challenges authors and empowers users by enabling them to interact with information. Moreover, tagging makes content accessible in different ways. As Tim O'Reilly states, the brain does not rely on "rigid categories" but creates "multiple, overlapping associations." This is apparent in the way that one person identifies and organizes information will not universally match the ideas of another person. An informal and open system of tags, comments, wikis, or other user-generated entries are collectively known as folksonomies, a term coined by Thomas Vander Wal. The root word, "folk," captures the democratic nature of this phenomenon. For, as Tom Steele states, "unlike a taxonomy, a folksonomy is created by those who are actually using the resources it classifies."⁹ Ellyssa Kroski similarly comments on this quality of folksonomies but as they relate to tagging: "Since a folksonomy arises as a result of user tagging, it is reflective of the way that they [users] categorize information."¹⁰ These reflections of a user's way of thinking not only capture thought processes of individual minds but also expressions of socio-cultural background, educational level, and economic standing, among other sociological attributes. A single, authoritative point of entry disregards diverse backgrounds in discovering and understanding information. Clay Shirky describes this problem in terms of communication theory by describing that act of merging multiple concepts into a single

⁹Steele, T. (2009). The new cooperative tagging. *Library Hi Tech*. 27(1), 68-77.

¹⁰Kroski, E. (2005, December 7). The hive mind: Folksonomies and user-based tagging. Retrieved from <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>

term as a “signal loss.”¹¹ Writers of the Library of Congress Classification Schedules partially anticipated this problem by offering alternative access points and directing users to the authority term with Used For, Broader Topic, Related Topic, See Also, and Narrower Topic notations. However, these directives still force the user toward a single, authoritative subject heading and access point for a topic. Tags, in contrast, do not prescribe meaning to a user but give expression to what users think they mean. Tagging removes some of the guesswork, or “mind reading and fortune telling,”¹² from creating access by placing the onus of developing adequate entry points on the users themselves.

Prevalence of Tagging

Given an understanding of what tagging can offer and mean for describing content, it seems pertinent to consider the extent of its popular adoption. In 2006, a study by the Pew Internet & American Life Project found that “28% of internet users have tagged or categorized online such as photos, news stories or blogs posts” and that “on a typical day online, 7% of internet users say they tag or categorize online content.”¹³ Of these users, demographic statistics show that taggers are considerably diverse in age, race, education, and economic income (see Figure 4).¹⁴

¹¹Shirky, C. (2005). Ontology is overrated: Categories, links, and tags. Retrieved from http://www.shirky.com/writings/ontology_overrated.html

¹²Ibid.

¹³Rainie, L. (2007, January 1). 28% of online Americans have used the Internet to tag content. *Pew Internet and American Life Project*. Retrieved from <http://www.pewinternet.org/Reports/2007/Tagging/Report.aspx>

¹⁴Demographics of taggers. (2007, January 1). [Table describing the demographics of taggers, based on a Pew Internet Poll tracking survey]. *Pew Internet and American Life Project*. Retrieved from <http://www.pewinternet.org/Reports/2007/Tagging/Report.aspx>

Demographics of Taggers	
<i>28% of online Americans say they have tagged content like a photo, a news story or a blog post</i>	
	Proportion of all Americans in the group who are taggers
Men	29%
Women	27%
Race/ethnicity	
White, non-Hispanic	26%
Black, non-Hispanic	36%
English-speaking Hispanic*	33%
Age	
18-29	32%
30-49	31%
50-64	23%
65+	18%
Educational attainment	
High school diploma	24%
Some college	28%
College degree +	31%
Household income	
<\$30K	28%
\$30K-\$49,999	28%
\$50K-\$74,999	27%
\$75,000+	36%
Internet connection at home	
Dial up	23%
Broadband	38%
Source: Pew Internet & American Life Project December 2006 tracking survey. N for internet users=1,623. Margin of error is $\pm 3\%$.	

Figure 4. Demographics of taggers, from PIP tracking survey.

One of the most popular sites for photo tagging is Flickr.com. As of April 2011, the Web traffic statistics site Alexa ranked Flickr #34 in the world and #24 in the United States for site hits.¹⁵ The same statistical summary stated that “visitors to the site view an average of 9.0 unique pages per day” and that they “spend about five minutes per visit to the site and 26 seconds per pageview.” This means that, for now, Flickr is a highly popular and actively used Web-based service. A related graph from 2007 shows steadily increasing site traffic (see Figure 2).

¹⁵ Alexa Flickr statistics.

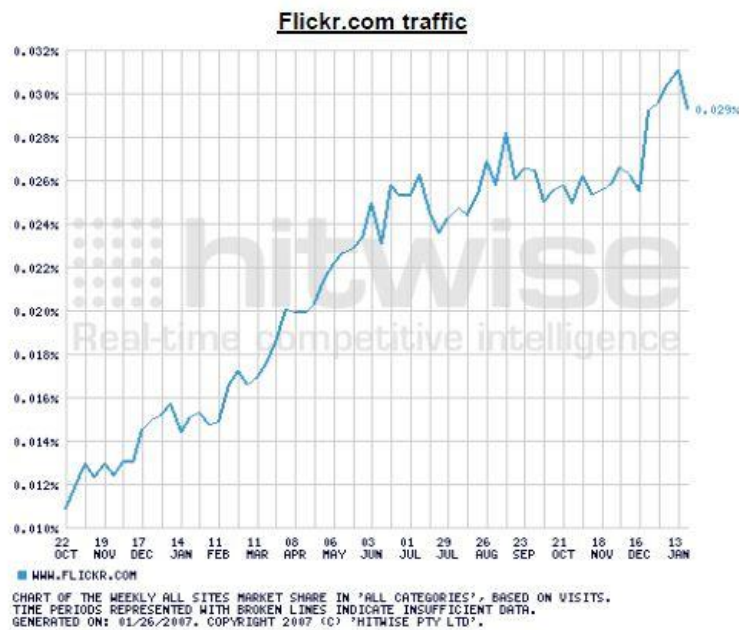


Figure 2. Graph of Flickr.com's climbing site traffic.

Current data¹⁶ from Alexa (see Figure 3) shows slightly dwindling use since 2007, possibly from competition with other social media sites with photo tagging functionality such as Facebook (ranked #2 in the world for site traffic), but Flickr's overall traffic ranking as #34 in the world places it in a secure position.



Figure 3. Alexa Flickr.com site traffic, a more recent graph.

¹⁶Flickr.com unique visitors.

In short, these data indicate considerable popularity of and engagement in social tagging. With such prevalence of this method for describing digital objects and information, it remains to know the extent that informational professionals, the traditional creators of descriptive authority and documentary access, have adopted this technology.

Professional Acceptance of Tagging

Informational professionals could understandably be perceive tagging as a threat, as it allows non-professional users to engage in creating access and description to resources, work traditionally done by cataloging and metadata specialists. However, a major study in 2005 endowed by the Institute for Museum and Library Services (IMLS), titled Steve: The Museum Tagging Project, found that, with some reservations aside, museum and archive professional seem to embrace tagging in conjunction with institutional access and retrieval systems. The Steve project was initially developed from a growing concern that users of digital collections struggle with searching for materials because of a “semantic gap” between professional terminology and user vernacular. The outcomes of this study showed that 86% of user tags “did not match existing museum metadata,” thereby revealing that a public language that reflected “the broad range of needs and perspectives of users, simply did not exist in collection documentation.”¹⁷ Returning to the postmodern criticism of monolithic narratives, project team members believed that tagging would add a “multi-cultural, perhaps multi-lingual perspective to museum documentation” and promote “strategies for engaging new types of users in

¹⁷Stein. R. (2009, January 1). Researching social tagging and folksonomy in art museums: Final report. *Institute of Museum and Library Sciences*. Retrieved from <http://verne.steve.museum/Steve.Museum-Final-Report.doc>

looking at and thinking about art.” When asked to evaluate the user-generated tags, museum professionals judged 88% of all tags “as being useful for finding or describing the particular works of art from their collections.” This percentage increased to 96.8% when tags were repeated several times by multiple users. In terms of professional attitudes toward tagging, the Steve project found that “approximately 75% of survey respondents agreed that museums could use social tagging for describing works of art or for aiding search and retrieval.”

Another study that also evaluated professional use of and attitudes toward social tagging found similarly supportive responses. Mary Samouelian surveyed 213 archival institutions “to determine the extent to which they are using the Web’s next generation of applications with respect to their digital collections.”¹⁸ Her content analysis showed that, of those 213 repositories, 85 included digital collections and 38 of those 85 (45%) repositories adopted some sort of Web 2.0 technology. In addition to finding a level of adoption, Samouelian also surveyed reasons for adopting or impeding integration of Web 2.0 technologies into digital collections. The top reason among professionals for supporting adoption was a desire for “increased promotion of department and resources,” with 57% of respondents saying so. The top reason for avoiding adoption was “time” (likely meaning money or resources), with 71% of respondents saying so, followed by 29% of respondents saying “lack of consistency with descriptive standards.” Samouelian concludes that “many archival professionals *are* embracing Web 2.0 [technologies] to promote their digital content and redefine their relationship with their patron.” So, while

¹⁸Samouelian, M. (2009). Embracing web 2.0: Archives and the newest generation of web applications. *American Archivist*, 72(1), 42-71.

adoption of tagging and other Web 2.0 applications is not unanimous, information professionals are not as resistant to adoption as one might imagine. As Guy & Tomkin state, the broad consensus among professionals for attitudes toward tagging could be summarized as tagging being “no replacement for formal systems” but capable of creating “systems that are conducive to searching, sorting and classifying.”¹⁹

Criticisms of Tagging

Despite this qualified broad acceptance of social tagging and Web 2.0, these new methods of discovering, arranging, and describing information are not without vocal critics. Perhaps most prominently, Andrew Keen disparages the non-professional, or “amateur,” contributor to information systems as “a digitalized version of Rousseau’s noble savage, representing the triumph of innocence of experience, of romanticism over the commonsense wisdom of the Enlightenment.”²⁰ Keen is not a mere elitist, and he does offer some valid criticisms of outsourcing information management to people untrained in best practice standards. Keen argues that the overwhelming volume of data and information produced electronically makes information professionals more valuable than ever, not obsolete or replaceable by the digital “noble savage.” He also notes how information professionals are ideally deeply familiar with available resources and trained to identify “what’s important and what’s not, what is credible from what is unreliable, what is worth spending our time on as opposed to the white noise that can be safely

¹⁹Guy, M., Tonkin, E. (2006, January). Tidying up tags? *D-Lib Magazine*. 12(1). doi:10.1045/january2006-guy

²⁰Keen, A. *The cult of the amateur*, 36.

ignored.”²¹ This implies that Keen believes that non-professionals, on the other hand, are for the most part either gullible or simply complacent when sifting through bad information.

Besides the potentially unreliable and inconsistent quality of non-professional work, as argued by Andrew Keen, David Weinberger argues that tags pose inherent contextual and semantic problems. Specifically, Weinberger asserts that “tags capture only a few bits ... because tags by themselves have no context” and that the miscellaneous quality of tags leads to “diminishing [of] their meaning and utility.”²² This seems like an irrelevant statement, as all meaning is arguably derived from relationships. Moreover, the purpose of this study, and its anticipated outcome, is to show what kind of context tags do, in practice, capture. Cal Lee summarizes the meaning of context as a “set of things, factors, or attributes that are related to a target entity in important ways but are not so closely related to the TE that they are considered to be part of the TE itself.”²³ He then proceeds to argue that, outside of a hypothetically omniscient agent, a full grasp of context is elusive, if not impossible, and that “there is no such thing as a digital monad, i.e. a fully self-contained, self-describing digital object that represents the entire universe.”²⁴ Thus tags, with their hyperlinks to digital objects and other associated tags, *build* a finite informational universe and create context by developing a network of semantic relationships.

²¹Ibid., 45.

²²Weinberger, D. *Everything is miscellaneous*, 165.

²³Lee, C. (2011). A framework for contextual information in digital collections. *Journal of Documentation*. 67(1), 95-143.

²⁴Ibid.

Besides his criticism of the potential loss of context, Weinberger finds tagging, or “the miscellaneous,” problematic because the act of organizing one thing results in the disordering of something else, an informational chaos theory ripple effect: “the basic fact that order often hides more than it reveals has sometimes itself been hidden within the art and science of organizing our world.”²⁵ Thus, for Weinberger and probably Keen, opening up organization and description to non-professional users is potentially chaos-inducing, an informational Pandora’s box. For this reason, Weinberger and Keen fiercely defend continuing the traditional practice of standardization through authority files and controlled vocabularies for the sake of a predictable process of search, access, and retrieval: “if you know that vocabulary (or if you browse the tree) you don’t have to guess.”²⁶ In this respect, Weinberger and Keen raise valid points. However, they also create a straw dog for their argument. Throughout the literature on tagging, as already noted, most promoters of Web 2.0 applications do not advocate tagging and other social networking tools as *replacements* for arrangement and description but as *supplements*. Yes, a tag cloud often represents a jungle of homonyms, synonyms, misspellings, and downright misleading terms, but most embracers of Web 2.0 are not proposing to eliminate the Library of Congress authority files or classification schedules.

Practical Problems with Tagging

Andrew Keen and David Weinberger mostly draw attention to theoretical or philosophical problems with Web 2.0, the “noble amateur,” and everything being “miscellaneous.” Some of their criticisms are valid and should be granted consideration.

²⁵Weinberger, D. *Everything is miscellaneous*, 88.

²⁶*Ibid.*, 90.

However, as the polling data demonstrated, Web 2.0 is here to stay, and the Web will likely only become more interactive as new technologies emerge. Given this state of affairs, it makes more sense to focus on practical criticisms of Web 2.0 and tagging functionality. Much of the literature on social tagging repeatedly identifies the following issues with social tags: “idiosyncrasy, inconsistency, contradiction, and inaccuracy” as well as “misspellings, poor encoding, acronyms, punctuation and compound tags that omit spaces.”²⁷²⁸ In communication theory terms, these problematic tags create “noise.” The lack of standards does not guarantee quality control and therefore permits various anomalies, disrupting the signal between a person’s intended meaning through tags and the recipients actual received message.²⁹ Hunter et al. and Heymann et. al. state that such errors may not pose a significant problem if a “critical mass” of taggers compensates for errors.³⁰³¹ However, in lower populated collections with few taggers, the lack of broad input increases the likeliness of problematic tags.³²

²⁷Hunter, J., Khan, I., Gerber, A. (2008). HarvANA: Harvesting community tags to enrich collection metadata. *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries*. New York, NY: ACM.

²⁸Mendes, L., Quiñonez-Skinner, J., Skaggs, D. (2009). Subjecting the catalog to tagging. *Library Hi Tech*. 27(1), 30-41.

²⁹Macgregor & McCulloch, Collaborative tagging as a knowledge organisation and resource discovery tool, 4.

³⁰Hunter et al. HarvANA: Harvesting community tags to enrich collection metadata, 149.

³¹Heymann, P., Koutrika, G., Garcia-Molina, H. (2008). Can social bookmarking improve web search? *WSDM '08 Proceedings of the international conference on Web search and web data mining*. New York, NY: ACM.

³²Hunter et al., HarvANA: Harvesting community tags to enrich collection metadata, 149.

Another criticism concerns a lack of recall, “the ability of a system to return all resources related to a topic.”³³ Without a hierarchy of standardized terms but instead a jumble of idiosyncratic tags, users of digital collections may find discovering a comprehensive set of relevant results a frustrating process. For example, the majority of images of cars may be associated with the tag “automobile” while a few separate images might only be associated with the plural of the same term, “automobiles.” A mere letter may determine the difference between partial and full retrieval of relevant digital objects. For this reason, Kroski believes that traditional taxonomies “provide a deeper, more robust classification of entities.”³⁴ For now, this is probably true. Tags would benefit from considerable clean-up so that users can find optimal search results. For this reason Kroski concludes that tags are best treated as “discovery systems,” not a precision-based research tool.

Lastly, one major criticism of tagging concerns the potential for limited relevance of tags to people besides the creator. Thomas Vander Wal characterizes folksonomies as being either “broad” or “narrow.”³⁵ A broad folksonomy offers a public service for many people interested in discovering objects relevant to their research, whereas narrow folksonomies are self-serving, with ambiguous or unknown meaning to outsiders. Guy & Tonkin share a similar criticism of tagging relevance: “Possibly the real problem with folksonomies is not their chaotic tags but that they are trying to serve two masters at

³³Kroski, The hive mind: Folksonomies and user-based tagging.

.

³⁴Kroski, The hive mind: Folksonomies and user-based tagging.

³⁵Vander Wal, T. (2005, February 21). Explaining and showing broad and narrow folksonomies. Retrieved from http://www.personalinfocloud.com/2005/02/explaining_and_.html

once; the personal collection, and the collective collection (12).”³⁶ An example of such a tag would be “to_read.” This narrow label tells an outsider nothing about the tagged object other than another user wants to read it at some point. Considering these obstacles and weaknesses, tagging is clearly not a perfect system. However, given sufficient monitoring, suggestion systems, a user registration requirement, or other methods of quality control, online communities could easily improve tags for more meaningful contributions to organizing information.

Tagging and Archival Processing

Having weighed the advantages and disadvantages of tags and shown that archivists do, overall, support the adoption of tagging and other Web 2.0 applications, it remains to discuss how tagging systems might assist professionals in responding to archival concerns. One of the most popular issues for user studies concerns access. The Society for American Archivists’ online glossary defines access as “the ability to locate relevant information through the use of catalogs, indexes, finding aids, or other tools.”³⁷ Since locating relevant information is achieved through access points, and access points are a part of archival processing, then users cannot locate relevant information without adequate processing. In Greene & Meissner’s seminal article, “More Product, Less Process,” they argue that adequate means finding a “golden minimum” where archivists

³⁶Guy & Tonkin. D-Lib Magazine, “Tidying up Tags?”

³⁷Access. (2005). In *A glossary of archival and records terminology*. Retrieved from http://www.archivists.org/glossary/term_details.asp?DefinitionKey=161

process collection enough to “maximize the accessibility of collection materials to users.”³⁸

To achieve a functional balance between process and access, Greene & Meissner urge archival professionals “to articulate a new set of arrangement, preservation and description guidelines” that decreases processing time, increases intellectual access, and increases use. Given limited time and resources, item-level description is an unrealistic, if not impossible, expectation. Steele states that a widespread problem among libraries, and by extension archives, is that these institutions accumulate materials faster than professionals can process them.³⁹ Particularly with the rapid proliferation of electronic resources, processing becomes a nearly Sisyphean task. In response to this challenge, Steele proposes that “the need for metadata can be alleviated by tagging.”⁴⁰ As it has already been stated, tagging would not be a replacement for professionally generated metadata and catalog entries, but adopting tagging as a means to crowd-source processing could minimize backlogs, create improved access, and ultimately engage users with previously hidden materials. As Kroski states, “in the absence of a professionally designed taxonomy, folksonomies are being viewed as a readily available, ‘better than nothing’, stand-in.”⁴¹ Tagging might not be the ultimate solution to expanding backlogs, but it does offer a pragmatic approach to creating better access. Ultimately, archivists do not need to view taggers as antagonistic to their profession status; instead, as Max Evans

³⁸Greene, M., Meissner, D. (2005). More product, less process: Revamping traditional archival processing. *The American Archivist*. 68(2), 208-263.

³⁹Steele, T. (2009). The new cooperative tagging. *Library Hi Tech*. 27(1), 68-77.

⁴⁰Ibid.

⁴¹Kroski, The hive mind: Folksonomies and user-based tagging.

asserts, “acting as partners with archivists, users can do what archivists alone cannot do.”⁴²

Context Entities of Tags: Context on Context

Having demonstrated the prevalence, strengths and weakness, and practical implications for archives of tagging, it now remains to discuss the contextual bearing tags bring to digital objects. As mentioned earlier, in response to Weinberger’s criticisms, documents gain enhanced relevance and meaning through connections with other documents.⁴³ In a tagging system, such relationships are built through hyperlinking terms to relevant items. When two or more objects are viewed together, the researcher builds context and gains a deeper understanding of the materials. Besides giving some context about the document, the tags themselves can fall into categories of what Cal Lee identifies as, “contextual entities.” These categories represent a type of metadata that contextualize context, allowing the user a sense of what *kind* of context is generated by user tags. It is interest this type of context that drives the core of this paper’s data collection and analysis.

Lee identifies nine classes of contextual entities “to elaborate the minimum number of categories of contextual entities that would be required to comprehensively document the “life history” of a target digital object.”⁴⁴ For the purpose of this study, I modified Lee’s classes to create ten context entities that describe the kind of information

⁴²Evans, M. (2007). Archives of the people, for the people, by the people. *The American Archivist*. 70(2), 387-400.

⁴³Lee, A framework for contextual information in digital collections, 10.

⁴⁴Lee, A framework for contextual information in digital collections, 10-11.

that tags capture (see methodology section). An understanding of the classes of informational context that tags capture could lead to improving tagging systems. Raya Fidel's article on automatic indexing argues that traditional indexing through content analysis and "translation" into key concepts (or union titles and subject headings) by professionals results in a "static" system. Fidel writes of a document containing "aboutness," basically its subject. This is a primary bit of information that a catalog entry must identify. However, beyond "aboutness," a document also captures "ofness," or pieces of information that do not directly relate to the subject of a document but express other information that is contextually relevant.⁴⁵ With this in mind, a document's "aboutness" remain relatively fixed in time, but its "ofness" or broader contextual relevance, will change according to circumstances—including the both the cataloger and the user's background, experience, and intended use of the object. With that said, dynamic systems such as tagging "are more promising as user-centered indexing methods because they may tailor indexing to the requirements of each request" as those requirements change with time and circumstances.⁴⁶ So, although in this article Fidel specifically discusses computer-automated indexing, her suggestion to improve intellectual indexing through adopting "dynamic" processes is relevant to user tagging, a characteristically adaptive way of describing digital objects. Given the multifaceted significance and use of context, it then remains to discover what kind of contextual information user tags tend to generate.

⁴⁵Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*. 6(3), 39-62.

⁴⁶Fidel, R. (1994). User-centered indexing. *Journal for the American Society for Information Science*. 45(8), 572-576.

Methodology

This study aims to quantify the frequency of different types of context entities assigned to tags for describing images. The image set analyzed in this study consists of photos uploaded to Flickr by university archives, public libraries, and museums. By determining the frequency of context entities, it is proposed that one will gain some understanding for what kind of information the tags on social photo sharing sites tend to identify and emphasize. Furthermore, the process of assigning these specific context entities to tags may reveal what, if any, contextual gaps among tags exist and, consequently, what classes of tags require more focus and expansion.

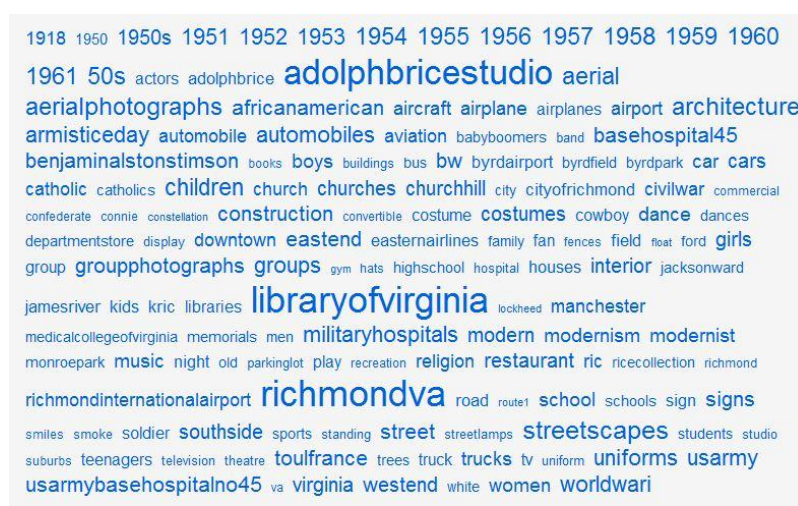
To gather the initial raw data, three university libraries, two large public libraries, and two museums were selected. These seven institutions were selected based on their participation in Flickr's The Commons, the "population" for this study, with all institutions uploading images from their collections to a public social media space open for assigning tags and posting comments. The stated objective of Flickr's The Commons is twofold: 1. "to increase access to publicly-held photography collections," and 2. "to provide a way for the general public to contribute information and knowledge (Then watch what happens when they do!)."⁴⁷

Participating institutions often included their own statement on usage rights and the purpose of uploading images to Flickr's site. A typical statement, such as the

⁴⁷Flickr: The commons. (2011). Retrieved from <http://www.flickr.com/commons/#faq>

following one from the George Eastman Kodak House, usually goes like so: “George Eastman House is participating in The Commons on Flickr to further the Museum’s educational mission and increase public access to its collection of photographs.”⁴⁸

Different institutions (university, public library, and museum) were selected for the sake of variety in material (institutional focus that may be limited to or expand beyond specific time periods, geographic locations, and specific cultures) and the types of users likely to comment (i.e.: academic researchers, amateur historians, casual browsers, genealogists, and curators, among others). After selecting these institutions, the next step involved pulling the data from associated Flickr pages and entering that into a spreadsheet. Visiting an institution’s Flickr page, one can click on a hyperlink to view the 150 most popular tags (see image below).⁴⁹



• This is a list of The Library of Virginia's 150 most popular tags. You can see [all The Library of Virginia's tags here](#).

Figure 4. Top 150 tags at the Library of Virginia’s Flickr site.

⁴⁸George Eastman House and The Commons on Flickr. (2011). Retrieved from <http://www.eastmanhouse.org/flickr/statement.php>

⁴⁹Flickr: The Library of Virginia’s 150 most popular tags. (2011). [Tag cloud of the top 150 user-generated tags for images on The Library of Virginia’s site on Flickr’s The Commons]. Retrieved from http://www.flickr.com/photos/library_of_virginia/tags/

One can also choose to view all of the tags then click on a specific tag to view associated images. This was often necessary because of problems with tag ambiguity. For example, the tag “2217” expresses no obvious meaning, but clicking on its hyperlink led me to an image of a train locomotive with the number “2217” on its front, a descriptor of the object. Since all of these tags were formatted with hyperlinks, which were problematic for the purpose of exporting to a spreadsheet, the data was first added into a batch editing program, TextSpesso. After batch editing the data to remove formatting, the find and replace function in Google Docs eliminated extra spaces and commas. With the data set converted to a simple, uniform stream of text, it was then imported into Microsoft Excel for better organization with a spreadsheet. At this point, I imported the data into a spreadsheet for organization and labeling context entities.

Assigning context entities to image tags involved a highly time-consuming process. A second column, next to the tags, was created to enter context entity labels, specifically (see Figure 5).

	A	B
5457	tourism	AB
5458	tourists	AG
5459	towers	OB
5460	town	PL
5461	townhalls	PL
5462	trademarks	AB
5463	traditionalmusic	F
5464	trafficsafety	AB
5465	trafficsigns	OB
5466	trail	OB
5467	train	OB
5468	translators	AG
5469	transportation	AB
5470	tredwellhouse	PL
5471	tree	OB

Figure 5. Spreadsheet sample of context entities assigned to tags pulled from Flickr.

As stated earlier, these context entities were adapted from an article by Cal Lee. Lee discusses issues concerning context of digital objects. Specifically, a user viewing a digital object viewed on its own might fail to grasp the broader significance of an object if it is removed from its socio-historical, authorial, geographic, temporal, and linguistic context, among other factors. As “the final arbiters of contexts,” human agents must fill in the contextual gaps. These contextual gaps are informational omissions created by the nature of the medium—the photograph distances the image from its original moment in time—then made more problematic by the process of digitizing an artifact now not only removed from time and space but also removed from its analog existence—existing in a digital realm, with possibilities for cropping, mash-ups, and countless other forms of modification that lessen authenticity. In response to this informational deficiency, tags or other forms of metadata restore a significant, if imperfect, degree of information about

the original context. The entities identified above distinguish useful classes for identifying the contextual nature of a tag.

Lee goes into greater detail about what he proposes to capture with each context entity, but for the purposes of this paper I will present a table of the tags (see Figure 6) as they were adapted for use in this study, with quoted definitions excerpted from Lee's article.

Object	A "bounded discrete entity" that "persists across multiple points in time and place." In some cases, an object referred to things larger than one might expect, such as a zeppelin or a tower. "Object" meant something without agency, replicable, and not specific enough to command a unique position in cultural memory or value; for instance, the Berlin Wall would be classified as "Place," not Object." Lastly, any non-human creature was designated an object. This was not to relegate animals to an exploited role but to acknowledge that, philosophically, most people would agree that animals do not possess the same degree of conscious agency as humans.
Place	A "designated point or region in space." This meant specificity—proper nouns. It also meant geospatial coordinates and large, natural, areas that exist without human artifice, such as a forest or a river.
Agent	An entity that can carry out actions." This included individuals, corporate bodies, groups, and various roles (i.e.: king, farmer, tattooist).
Abstraction	A catch-all class for things that were not an agent, object, or place—"properties or qualities as distinguished from any particular embodiment." Some examples included: activities or fields (i.e.: walking, aviation, politics), ethnicity or races (i.e.: Albanian, American), adjectives (i.e.: sunny, somber), or anything else conceptual and not specific to a form of expression or being.
Occurrence	A momentous event or activity, such as the Columbian Exposition or the Civil War, but also events like a wedding or an inauguration.
Time	Minutes, hours, months, seasons, years, centuries, eras, and so forth.
Form	A "particular way of expressing ideas or information," such as: daguerreotype, opera, novel, sculpture, and other manifestations of genre or expression.

Foreign Lang.	Assigned only to non-Roman alphabet based languages, such as: Japanese, Chinese, Russian, Arabic, and Hebrew. European languages were translated and then assigned an appropriate entity.
Identifier	DOIs, catalog numbers, and other labels, usually numbers, applied for machine-readable functionality
Uncertain	Tags that were so arcane, ambiguous, poorly spelled, or personal that examining the image or searching for clarification online could not help explain the intended meaning.

Figure 6. Context entities as adapted for this study.

When using Lee's context entities, I needed to make some modifications. For example, I did not use the "relationship" entity, defined by Lee as "an association between two or more entities (or classes of entities), which cannot be reduced to or adequately expressed as a property of the entities (or classes of entities) themselves." I did not use this entity because I only assigned one entity per tag, and the relationship entity depends on assigning a tag multiple entities. In retrospect, it would have been beneficial to assign each tag multiple entities, if applicable. Not doing so made assigning entities more arbitrary in that I needed to judge the most representative "ofness" of a tag at the expense of other, still relevant, entities. Being a subjective decision, the assigned entity would vary across coders. This issue is addressed further in the "discussion" section of this paper. Also, I did not use the "Purpose" entity, defined by Lee as "mandate, norms, values, intentions, rules, standards, virtues, or functions to which agents can advance or with which they can conform." I thought that guessing a tagger's intended purpose, as defined so, was beyond my ability and would end up being "mind reading or fortune telling," as I earlier quoted Shirky. Additionally, I divided "Form of Expression" into "Form" and "Foreign Language." I did so for two reasons. First, I interpreted Form of Expression as signifying a medium or genre, not a specific language. However, the tags often did represent foreign languages, some of which I was unable to

translate, so I decided that it would be better to make a separate note of these instances rather than lumping them in the “uncertain” entity. Finally, I created an “Identifier” entity to account for many computer-generated tags. As automated tags, one cannot properly describe them as the results of social tagging. As it will be acknowledged in the findings section, ID entities represented a disproportionately high percentage of the total tags. For a better sense of the proportion of other, human-generated, entities, it proved useful to exclude the ID entity during my analysis.

By looking at the tag alone, the appropriate entity would seem self-evident. In practice, identifying a tag’s entity often posed problems with ambiguity. This required following the hyperlink to associated images and viewing the tag’s context to clarify the tagger’s possible intention. As already stated, in some cases the tag seemed so irrelevant or cryptic that a “U” or “Uncertain” entity was needed to acknowledge ambiguity or confusion. For example, one image from the Eastman House shows a contortionist bending over backwards. A user assigned the tag “hooo,” which expressed no clear meaning for me, and likely other people. Another example is the tag “flt,” which might abbreviate a full word for someone, but it is not obvious for everyone else. The “U” entity was also used to describe obvious instances of spelling errors or typos that were otherwise entered as a properly typed separate tag. This approach eliminated unintentional redundancies that were distinct from intentional redundancies, such as creating a tag for the same word but in a different language.

Findings

After adding the number of tags for each entity and the total number of tags per institution, it was shown that “identification” context entities represented the most number of tags (see Figure 6).

	AB	AG	F	FL	ID	OB	OC	PL	T	U	TOTAL
Field	107	123	6	0	2	510	33	320	29	17	1147
Eastman	714	715	61	17	1603	1287	88	442	96	56	5079
Michigan	128	128	7	0	0	100	53	123	5	1	545
UW	279	442	18	0	0	582	79	514	15	3	1932
Cornell	337	579	47	0	3732	1152	99	1362	16	27	7351
DC	73	154	0	0	3	140	20	158	19	10	577
NYPL	422	599	32	172	2643	874	61	1026	59	22	5910
TOTAL	2060	2740	171	189	7983	4645	433	3945	239	136	22541

Figure 6. Context entity totals by institution.

However, the ID entity was designated primarily for the sake of removing such tags from the overall entity pool. In most cases, these tags are created by the institutions, not users, so one cannot properly describe them as social tagging; there was no collaborative or user-input process involved in the creation of ID tags. In all likelihood, the institutions automated ID tag generation while uploading the image files to Flickr. The results are more representative of actual user tagging behavior if one excludes the ID tags (see Figure 7).

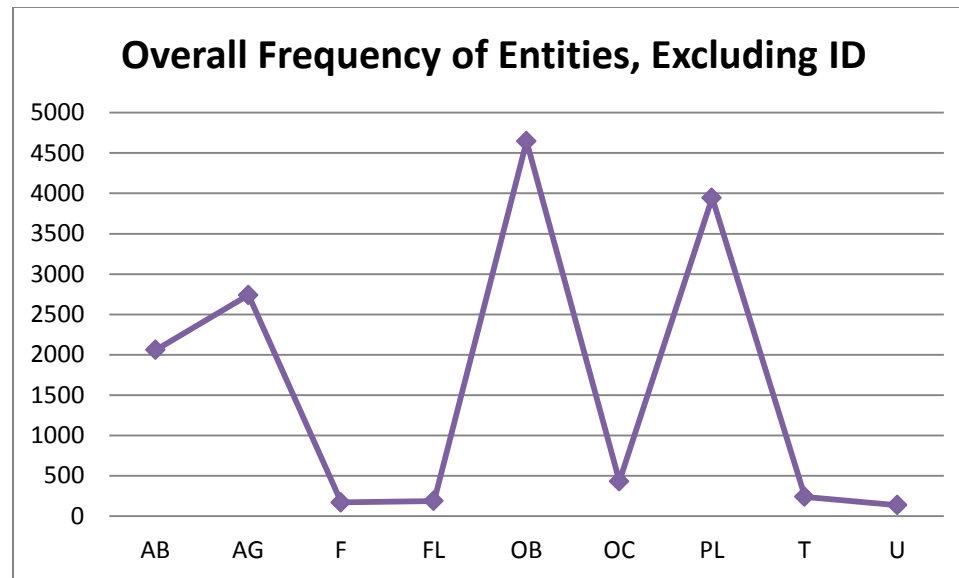


Figure 7. Frequency of total tag entities, excluding “ID.”

In this case, “objects” represented the most tags, with 32% of all tags captured by the OB context entity. The next highest percentage of tags was represented by the “place” entity, with 27% of the total tags being so. Next came “agent” (19%), “abstraction” (14%), “occurrence” (3%), “time” (2%), and “form,” “foreign language,” and “uncertain” all representing 1% of the total tags. That margin between the top two entities, OB and PL, and the third most frequent entity, AG, is considerably wide, at 8%, suggesting that the type of context given to digital images through tagging is dominated by descriptions of places and objects. This is probably not surprising. By looking at an image alone, removed from any background information, users will likely not be able to say much about an image other than what it physically represents. For example, a person may see an image of a plant that is a hybrid species and identify it as a “plant,” an object (OB), but that person will be less likely to tag the same image “hybrid species,” an abstraction (AB), by looking at the image alone. The same rule likely applies for the time and

occurrence entities. In other words, an entity more closely associated with specificity in time or concept—abstractions—will probably be represented less by tags.

These findings imply that, if users and professionals are to cooperate in describing collections, they should focus on different areas. Specifically, the employees at the institutions that originally post an image presumably have greater knowledge of provenance. This means that professionals must identify relevant agents, times, occurrences, places, abstractions, or any other entity that is not immediately apparent (or missing) from the image itself. Without access to full descriptions of provenance or other relevant background information, users can capture the “aboutness,” the subject, of an image much better than its “ofness,” or extended information not embedded in the image itself. From the raw data table (Figure 6) and its associated chart (Figure 7), one can see that Time, Form, and Occurrence are rarely captured by users. One could potentially discover pertinent information for these entities in provenance or donor agreement files—documents not typically paired with online images. So, while users can provide a service to their fellow users by creating a contextual web, or folksonomy, their areas of strength are limited. Thus, professionals, if they are to engage the public with their institutional collections by posting documents online, must fill the gaps in capturing entities.

Other graphs were created to give visual expression to and balanced consideration of the weight of each institution’s influence on entity percentages. For example, the Eastman museum, Cornell University, and the New York Public Library all uploaded a considerable amount more images to The Commons on Flickr than other institutions. The chart “Frequency of Entities per Institution” (see Figure 8) showed that the proportion of tag entities remained stable across institutions, with the exception of the Field and

Eastman museums and the New York Public Library, where these three institutions created a large number of tags associated with ID entities. Again, the ID entities were problematic because they were most likely the product of institutional professionals or machines, not tagging by non-professional users. Some remarkable anomalies appeared, though. At the University of Michigan, the percentage of “abstraction” entities was nearly equal to the percent of “agent” entities, and both of these entities passed “object” and “place” in percentage rank. Also, at the Washington, D.C., Public Library the percentage of “agent” entities was nearly comparable to the percentage of “place” entities and exceeded the percentage of “object” entities. These two anomalies are likely based on collection emphases of the institutions. For instance, the D.C. Public Library’s location in the political center of the United States probably resulted in a greater number of images of people, namely politicians and other national historical figures.

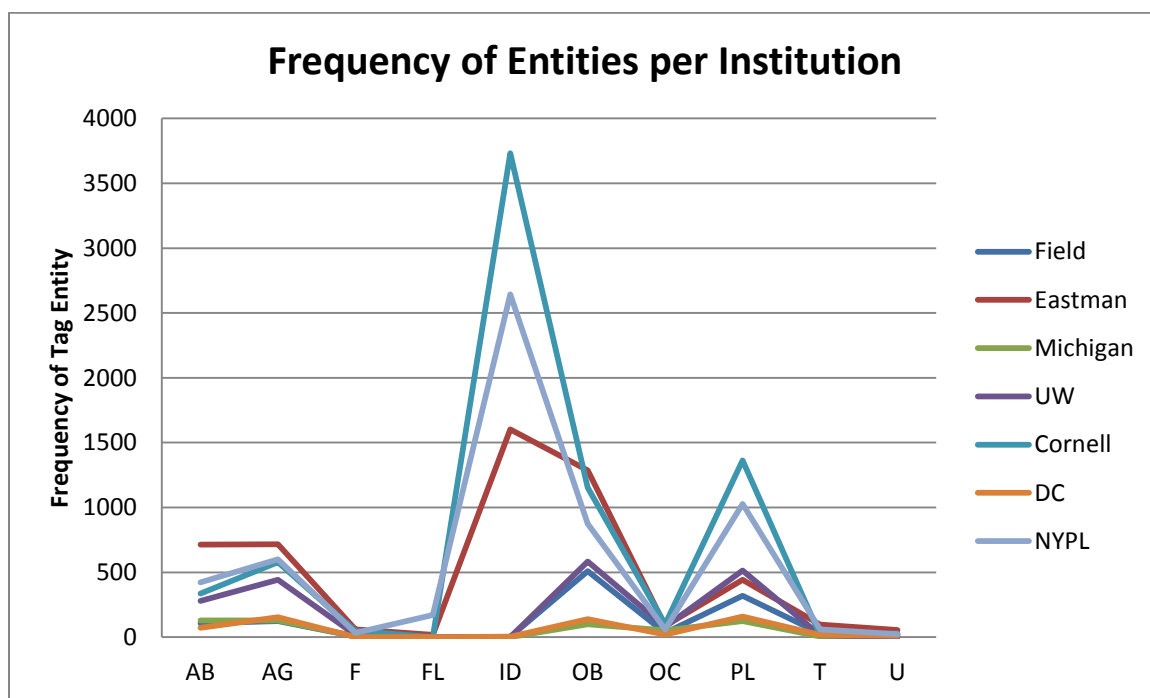


Figure 8. Frequency of context entities by institution.

Limitations and Suggestions for Further Research

Collectively, these data on entities suggest that tags on Flickr, and likely other social tagging sites for images, capture limited classes of information on digital images. As noted, tags tend to represent information about people, places, and things. This is not surprising, as those same categories are the most commonly reflected entities of our daily language; most people do not talk about abstractions, form, or genre, in casual conversation. As much of the literature states, tags offer a democratic method of describing information and, as one would expect, they are expressed through a community's vernacular, not jargon or conceptual language. Furthermore, archivists are privileged with special knowledge in some areas of context. For instance, by owning the original document, they might have access to more information about time, form, location, and occurrence. Eddie Adams' "Vietnam Execution" photo is familiar to many people. However, other war images might be unfamiliar to the majority of people who do not have access to information on provenance. Returning to Greene & Meissner's "More Product, Less Process" article, this means that archivists should focus on sharing information that only they can provide and letting users fill in the gaps when possible. This would be achieving that "golden minimum." Overall, then, these data show that institutions can probably depend on users to create alternative contextual terms for agents, objects, and places. However, social media sites such as Flickr could stand to increase the extent of information captured by other context entities such as time, occurrence, abstraction, and form. Archivists and taggers can and should cooperate to

build a more complete sense of a digital object's context and thereby augment meaning of collections.

Given an opportunity to replicate this data collection and analysis, this study could benefit from some helpful changes. First, just as a single term cannot capture the full range of possibilities for access, a single context entity assigned to each tag was not always adequate. In some cases where a tag could represent more than one entity, a choice of the most representative entity was made. A single entity is not always adequate for capturing the broader context of a tag. Instead, one could assign each tag multiple entities when relevant and thereby gather a broader representation of the dynamic contextual possibilities of tagging. My decision to assign only one entity per tag compounded the problem of subjectivity. To start, tagging is a highly subjective activity. With only one person subjectively classifying something that is already subjective, the full extent of the information captured by tags could be lost. Not only would it be better to assign multiple entities, when applicable, but recruiting multiple people to code the tags would create a higher degree of representativeness. Admittedly, I reproduced the very problem that critics of controlled vocabularies identify: single-person input for creating meaning. By conducting this study with multiple coders, I could gather interesting data about inter-coder reliability and ensure that pertinent entities are identified if another coder misses it. Second, and in relation to the first consideration for improvement, it would be worthwhile to mark tags in all foreign languages, not only non-Roman alphabet languages, with the FL entity. As already suggested, this could be achieved by assigning multiple entities to a single tag. More could be discovered about the diversity of language in tagging systems if all foreign languages were identified with

that entity. Finally, context entities would reveal more about tagging for the individual institutions if more image collections were uploaded to the Flickr site. Images in The Commons do not capture the full extent of an institution's digital collections; often only highlights or particular collections are shared. Perhaps by focusing on a single institution's full digital image collection and analyzing its proprietary tagging system, not Flickr, would reveal more meaningful statistics about tags and the type of context given to digital images through social tagging.

Conclusion

This study suggests that, despite general professional acceptance and a considerable extent of user participation, social tagging has definite limitations when considering the breadth of the entities they capture. Users create tags for entities that are more apparent in the image itself (object and places), but not as much so for entities that are not embedded in the image and require access to background information, such as a provenance file. In response, information professionals should be aware of contextual shortcoming and post background documentation or file information when the images themselves cannot impart enough information. If information professionals and users wish for the quality tagging to improve, then professionals should not simply post images without any explanatory context and then expect taggers to create useful and accurate tags on their own.

Contributors to the literature related to Web 2.0 applications and tagging functionality consistently identify several problems and limitations. However, by continuously reviewing what tagging gets right and what it misses, its value to creating and sharing information can be improved. Tagging holds much promise as a cooperative endeavor between professionals and users. Professionals can benefit from tagging by users creating access points that will help process collections otherwise hidden in a backlog. Users can benefit from professionals sharing institutional documentation related to image collections that would encourage a broader and more representative set of

context entities presented through tags. Tagging will continue to evolve into more precise and sophisticated systems, which will benefit any information seeker—amateur or professional.

Works Cited

- Access. (2005). In *A glossary of archival and records terminology*. Retrieved from http://www.archivists.org/glossary/term_details.asp?DefinitionKey=161
- Cook, T. (2001). Archival science and postmodernism: New formulations for old concepts. *Archival Science*, 1(1), 3-24.
- Cook, T. & Schwartz, J. (2002). Archives, records, and power: From (postmodern) theory to (archival) performance. *Archival Science*. 2(3-4), 171-185.
- Demographics of taggers. (2007, January 1). [Table describing the demographics of taggers, based on a Pew Internet Poll tracking survey]. Pew Internet and American Life Project. Retrieved from <http://www.pewinternet.org/Reports/2007/Tagging/Report.aspx>
- Evans, M. (2007). Archives of the people, for the people, by the people. *The American Archivist*. 70(2), 387-400.
- Fidel, R. (1994). User-centered indexing. *Journal for the American Society for Information Science*. 45(8), 572-576.
- Flickr: The commons. (2011). Retrieved from <http://www.flickr.com/commons/#faq>
- Flickr: The Library of Virginia's 150 most popular tags. (2011). [Tag cloud of the top 150 user-generated tags for images on The Library of Virginia's site on Flickr's The Commons]. Retrieved from http://www.flickr.com/photos/library_of_virginia/tags/

- George Eastman House and The Commons on Flickr. (2011). Retrieved from <http://www.eastmanhouse.org/flickr/statement.php>
- Greene, M., Meissner, D. (2005). More product, less process: Revamping traditional archival processing. *The American Archivist*. 68(2), 208-263.
- Guy, M., Tonkin, E. (2006, January). Tidying up tags? *D-Lib Magazine*. 12(1).
doi:10.1045/january2006-guy
- Heymann, P., Koutrika, G., Garcia-Molina, H. (2008). Can social bookmarking improve web search? *WSDM '08 Proceedings of the international conference on Web search and web data mining*. New York, NY: ACM.
- Hunter, J., Khan, I., Gerber, A. (2008). HarvANA: Harvesting community tags to enrich collection metadata. *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries*. New York, NY: ACM.
- Hyry, T. & Light, M. (2002). Colophons and annotations. *American Archivist*. 65(2), 216-230.
- Keen, A. (2007). *The cult of the amateur: How today's internet is killing our culture*. New York, NY: Doubleday/Currency.
- Kroski, E. (2005, December 7). The hive mind: Folksonomies and user-based tagging. Retrieved from <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>
- Lee, C. (2011). A framework for contextual information in digital collections. *Journal of Documenation*. 67(1), 95-143.
- Macgregor, G., McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5), 291-300.

- Matusiak, K.K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services: International Digital Library Perspectives*. 22(4), 283-298.
- Mendes, L., Quiñonez-Skinner, J., Skaggs, D. (2009). Subjecting the catalog to tagging. *Library Hi Tech*. 27(1), 30-41.
- O'Reilly, T. (2005, September 30). What is web 2.0: Design patterns and business models for the next generation of software. Retrieved from <http://owl.english.purdue.edu/owl/resource/560/10/>
- Rainie, L. (2007, January 1). 28% of online Americans have used the Internet to tag content. *Pew Internet and American Life Project*. Retrieved from <http://www.pewinternet.org/Reports/2007/Tagging/Report.aspx>
- Samouelian, M. (2009). Embracing web 2.0: Archives and the newest generation of web applications. *American Archivist*, 72(1), 42-71.
- Shirky, C. (2005). Ontology is overrated: Categories, links, and tags. Retrieved from http://www.shirky.com/writings/ontology_overrated.html
- Steele, T. (2009). The new cooperative tagging. *Library Hi Tech*. 27(1), 68-77.
- Stein, R. (2009, January 1). Researching social tagging and folksonomy in art museums: Final report. *Institute of Museum and Library Sciences*. Retrieved from <http://verne.steve.museum/Steve.Museum-Final-Report.doc>
- Vander Wal, T. (2005, February 21). Explaining and showing broad and narrow folksonomies. Retrieved from http://www.personalinfocloud.com/2005/02/explaining_and_.html
- Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital*

disorder. New York, NY: Times Books.