DHQ: Digital Humanities Quarterly
2015
Volume 9 Number 1

“By the People, For the People”: Assessing the Value of Crowdsourced, User-Generated Metadata

Christina Manzo <cmanzo_at_bpl_dot_org>, Simmons College, USA
Geoff Kaufman <geoff_dot_kaufman_at_dartmouth_dot_edu>, Tiltfactor Laboratory, Dartmouth College, USA
Sukdith Punjasthitkul <sukdith_dot_punjasthitkul_at_dartmouth_dot_edu>, Tiltfactor Laboratory, Dartmouth College, USA
Mary Flanagan <mary_dot_flanagan_at_dartmouth_dot_edu>, Tiltfactor Laboratory, Dartmouth College, USA

Abstract

With the growing volume of user-generated classification systems arising from media tagging-based platforms (such as Flickr and Tumblr) and the advent of new crowdsourcing platforms for cultural heritage collections, determining the value and usability of crowdsourced, “folksonomic,” or user-generated, “freely chosen keywords” [21st Century Lexicon] for libraries, museums and other cultural heritage organizations becomes increasingly essential. The present study builds on prior work investigating the value and accuracy of folksonomies by: (1) demonstrating the benefit of user-generated “tags” - or unregulated keywords typically meant for personal organizational purposes - for facilitating item retrieval and (2) assessing the accuracy of descriptive metadata generated via a game-based crowdsourcing application. In this study, participants (N = 16) were first tasked with finding a set of five images using a search index containing either a combination of folksonomic and controlled vocabulary metadata or only controlled vocabulary metadata. Data analysis revealed that participants in the folksonomic and controlled vocabulary search condition were, on average, six times faster to search for each image (*M* = 25.08 secs) compared to participants searching with access only to controlled vocabulary metadata (*M* = 154.1 secs), and successfully retrieved significantly more items overall. Following this search task, all participants were asked to provide descriptive metadata for nine digital objects by playing three separate single-player tagging games. Analysis showed that 88% of participant-provided tags were judged to be accurate, and that both tagging patterns and accuracy levels did not significantly differ between groups of professional librarians and participants outside of the Library Science field. These findings illustrate the value of folksonomies for enhancing item “findability,” or the ease with which a patron can access materials, and the ability of librarians and general users alike to contribute valid, meaningful metadata. This could significantly impact the way libraries and other cultural heritage organizations conceptualize the tasks of searching and classification.

1. Introduction

Classification is a basic, integral and historically significant human function. Defined by Golder and Huberman as an act “through which meaning emerges” [Golder and Huberman 2006, 200], the practice of classification represents one of the primary intellectual foundations of library and information sciences. Useful classification systems both accurately reflect the contents of a particular collection and allow for the effective and efficient retrieval of items or information. Given the subjective nature of human classification, examining meaning has no universal procedure. From the birth of the library science field until the late 1800s, during which the first edition of Dewey Decimal Classification was published in 1876, followed quickly by the Cutter Expansive Classification System and the introduction of Library of Congress Subject Headings [Stone 2000], all libraries and centers of information each used an independent, unstandardized systems of organization, otherwise known as “local classification.” 1

With the rise of such classifications as Dewey, Cutter and LC Classification, local classification was mostly retired in the late 19th century in favor of a united system that would allow for understanding across all libraries. This was, of course, 2

until the rise of Internet tagging-based platforms such as Flickr, Twitter, Tumblr, and Delicious presented a challenge to these standardized, centralized systems. Currently information professionals are facing an unprecedented amount of unstructured classification. Classification systems generated by such tagging-based platforms are referred to as "folksonomy," or "a type of classification system for online content, created by an individual user who tags information with freely chosen keywords; also, the cooperation of a group of people to create such a classification system" [21st Century Lexicon].

3 While folksonomies represented an increased diversity of classification, they were perceived mostly as sources of entertainment and documentations of casual colloquialisms rather than a formal system of documentation. However, in 2006 Golder and Huberman conducted a study of folksonomic data generated by users of the website Delicious demonstrating that user-provided tags not only formed in coherent groups, but also accurately described the basic elements of items that were tagged, such as the "Who," "What," "Where," "When," "Why," and "How" of items, proving their ability to supplement formal records.

4 Golder and Huberman's analysis also revealed two highly subjective categories that may diminish the potential value of crowdsourced metadata: "Self Reference" (i.e., any tag including the word "my," such as "mydocuments") and "Task Organizing" (i.e., tags denoting actions such as "readlater" or "sendtomom"). In examining the overall accuracy and reliability of tags, Golder and Huberman concluded: "Because stable patterns emerge in tag proportions, minority opinions can coexist alongside extremely popular ones without disrupting the nearly stable consensus choices made by many users" [Golder and Huberman 2006, 199].

5 This research was expanded upon in 2007, when Noll and Meinel compared tags from websites Delicious and Google against author-created metadata and found that the former provided a more accurate representation of items' overall content. Bischoff et al. [2009] also examined folksonomic metadata within the context of the music industry and found that tags submitted by users at the website Last.fm were comparable with professional metadata being produced at AllMusic.com. Syn and Spring [2009] examined folksonomic classifications within the domain of academic publishing, and also found authoritative metadata to be lacking when compared with its user-generated counterparts. As stated by Noll and Mienel, "tags provide additional information which is not directly contained within a document. We therefore argue that integrating tagging information can help to improve or augment document classification and retrieval techniques and is worth further research." [Noll and Meinel 2007, 186]. Together, these studies indicate that folksonomies can easily and usefully be stored alongside classical, controlled vocabularies.

6 While some may have reservations regarding the "mixing" of folksonomic and controlled vocabularies, these two classification systems need not be viewed as mutually exclusive. Both systems have inherent advantages as well as potential flaws. Controlled vocabularies are reliable and logically structured, but can be somewhat inaccessible to the casual user [Maggio et al. 2009]. Furthermore, they can be time-consuming and expensive to produce and maintain. For example, if one's local library employs two catalogers at an average salary of $58,960/year [American Library Association - Allied Professional Association 2008], it sets aside $117,920 for implementing controlled vocabularies on a limited number of new collection items alone. In contrast, folksonomies represent a relatively quicker and more cost-effective alternative [Syn and Spring 2009], with greater public appeal and accessibility, as evidenced by their overwhelming usage on social media. Mai noted: "Folksonomies have come about in part in reaction to the perception that many classificatory structures represent an outdated worldview, in part from the belief that since there is no single purpose, goal or activity that unifies the universe of knowledge" [Mai 2011, 115]. As many users have become accustomed to the level of service and interaction styles offered by current popular search engines, traditional searches are "unlikely to be very successful" and are becoming "less frequent as patrons' behavior is shaped by keyword search engines" [Antell and Huang 2008, 76]. As researcher Heather Hedden points out, if an artifact "[in today's culture] if it cannot be found, it may as well not exist" [Hedden 2008, 1]. However, their lack of centralization renders folksonomies prone to issues of potential data contamination, such as an unorganized, unstructured plurality of subjects and the likelihood of data duplication between users. Regardless, folksonomies and traditional systems of organizations may be used in tandem to address the shortcomings of their respective features, allowing for a more diverse and organized form of classification.

Our research aimed to provide new empirical evidence supporting the value of folksonomies by: (1) directly testing the benefits of adding user-generated folksonomic metadata to a search index and (2) comparing the range and accuracy of tags produced by library and information science professionals and non-professional users. The three main questions guiding this work were: 7

- RQ1: Will users exhibit reduced search times and greater hit rates when retrieving images with a search index that includes folksonomic metadata contributed by previous users?
- RQ2: Will general users and information science professionals differ in the type and quality of metadata they provide in a free-form tagging game?
- RQ3: Will users only provide information useful to them (e.g., Self Reference and Task Organization tags), or will they attempt to provide metadata that is useful on a wider scale?

2. Background and Overview of Present Research

The present research employed a hybrid form of usability testing utilizing the Metadata Games platform [http://www.metadatagames.org], an online, free and open-source suite of games supported by the National Endowment for the Humanities and developed by Dartmouth College's Tiltfactor Laboratory. The Metadata Games Project, launched in January 2014, aims to use games to help augment digital records by collecting metadata on image, audio, and film/video artifacts through gameplay [Flanagan et al. 2013]. Current Metadata Games media content partners include Dartmouth College's Rauner Special Collections Library, the British Library, the Boston Public Library, the Sterling and Francine Clark Art Institute Library, UCLA, and Clemson University's Open Parks Network. Inspired by other successful crowdsourcing efforts, the designers endeavored to create a diverse suite of games that could enable the public to engage with cultural heritage institutions and their digital collections, invite them to contribute knowledge to those collections, and set the stage for users to create and discover new connections among material within and across collections. The Metadata Games platform currently includes a palette of games that cater to a variety of player interests, including both single-player and multi-player games, competitive and cooperative games, and real-time and turn-based games, available for browsers and/or mobile devices. Despite their variety, all the games in the suite are united by a common purpose: to allow players to access media items from a number of cultural heritage institutions' collections and provide them with the opportunity to contribute new metadata, in the form of single-word or short-phrase tags, within the context of an immersive, enjoyable game experience. The end result is that institutions benefit from increased engagement from a variety of users and acquire a wider array of data about their media collections. 8

This research represents a collaborative project between the first author, who chose to use the Metadata Games platform as the focus for an independent study project at the Simmons College of Library and Information Sciences, and the co-authors from Dartmouth College's Tiltfactor Laboratory. To be clear, the goal of the reported study was not to provide a validation of the Metadata Games platform, but rather to study the value of folksonomic metadata more generally; that is, the focus of this research was on the data itself, and the tool employed was intended to be largely incidental and peripheral to the study's aims. At the time, the Metadata Games project was one of the few open-source metadata gathering tools available for cultural heritage institutions.[1] Thus, while the reported study results *are* specific to datasets gathered by the Metadata Games platform, the conclusions drawn from this study are intended to be generalizable to any organization currently making use of services such as CrowdAsk, LibraryThing or Scripto or considering a crowdsourced metadata project. 9

3. Methods

According to Nielsen, the number of participants needed for a usability test to be considered statistically relevant is five [Nielsen 2012b]; however, because of the additional collection of quantitative data in our hybrid study, sixteen individuals (eight men and eight women; six of whom were aged 18-24, eight aged 25-44, one aged 45-60, and one over 60 years of age) were recruited to participate individually in 30-40 minute sessions. In order to discern any differential patterns of results between librarians and non-librarians, and to separate out any potential advantage that users in the field of Library and Information Sciences might have with content search and metadata generation, a mixed sample (with nine participants recruited from LIS-related fields and seven from non-LIS-related fields) was used for the 10

study.

The study was divided into two main tasks. In the first task, participants were presented with physical facsimiles of five images from the Leslie Jones Collection of the Boston Public Library and instructed to retrieve these items using the Metadata Games search platform. The images presented to participants were divided into the following categories: Places, People (Recognizable), People (Unidentified or Unrecognizable), Events/Actions, Miscellaneous Formats (Posters, Drawings, Manuscripts etc.), as seen in Figures 1-5 below. Upon being given each physical facsimile, participants were timed from the moment they entered their first search term until the moment they clicked on the correct digital item retrieved from the Metadata Games search platform. This practice was adapted to reflect the digital items that would most commonly exist in a typical humanities-based collection, (i.e., photographs, manuscripts, postcards, glass plate negatives and other miscellanea). This design mirrored the common everyday occurrence of patrons attempting to retrieve a specific media item that they have in mind when using a library search index. According to a 2013 PEW Research Study, 82% of people that used the library in the last 12 months did so looking for a specific book, DVD or other resource [PEW Internet 2013]. 11

For this image search component of the study, participants were randomly assigned to one of two search index conditions: one with access to controlled vocabularies and folksonomic metadata (i.e., the “folksonomy condition”) and the other with restricted access only to controlled vocabularies (i.e., the “controlled vocabulary condition”) [See Figure 6 for a schematic representation of the study design]. Searches were conducted using two different, customized versions of the search index on the Metadata Games website [play.metadatagames.org/search]. The folksonomic metadata was generated by previous users of the Metadata Games platform, whereas the controlled vocabularies attached to the items were generated by Boston Public Library staff. The process of inputting the controlled vocabularies into both versions of the search index required some reformatting. For example, because the version of the search platform used in the study did not allow for special characters such as the dash “-” or the comma “,”, terms such as “Boston Red Sox (baseball team)” were imported as two individual tags: “Boston Red Sox” and “baseball team.” 12

Figure 1. Search Image 1; Category: Places.

Figure 2. Search Image 2; Category: People (Recognizable).

Figure 3. Search Image 3; Category: People (Unrecognizable).

Figure 4. Search Image 4; Category: Events/Actions.

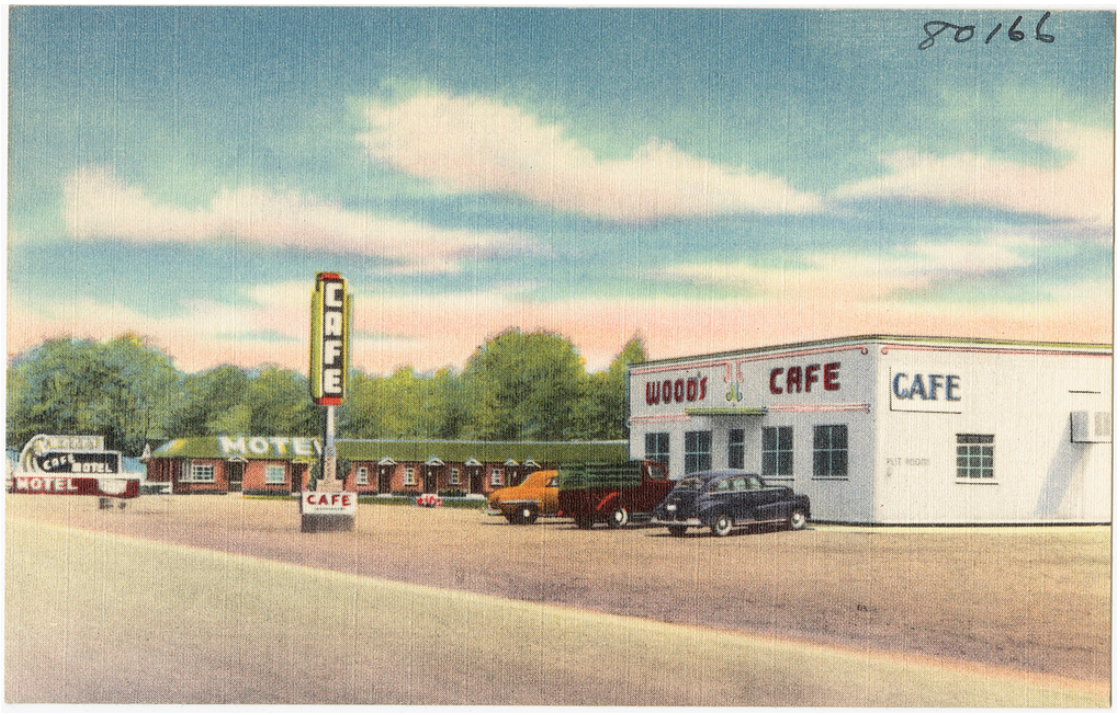

Figure 5. Search Image 5; Category: Misc. Formats.

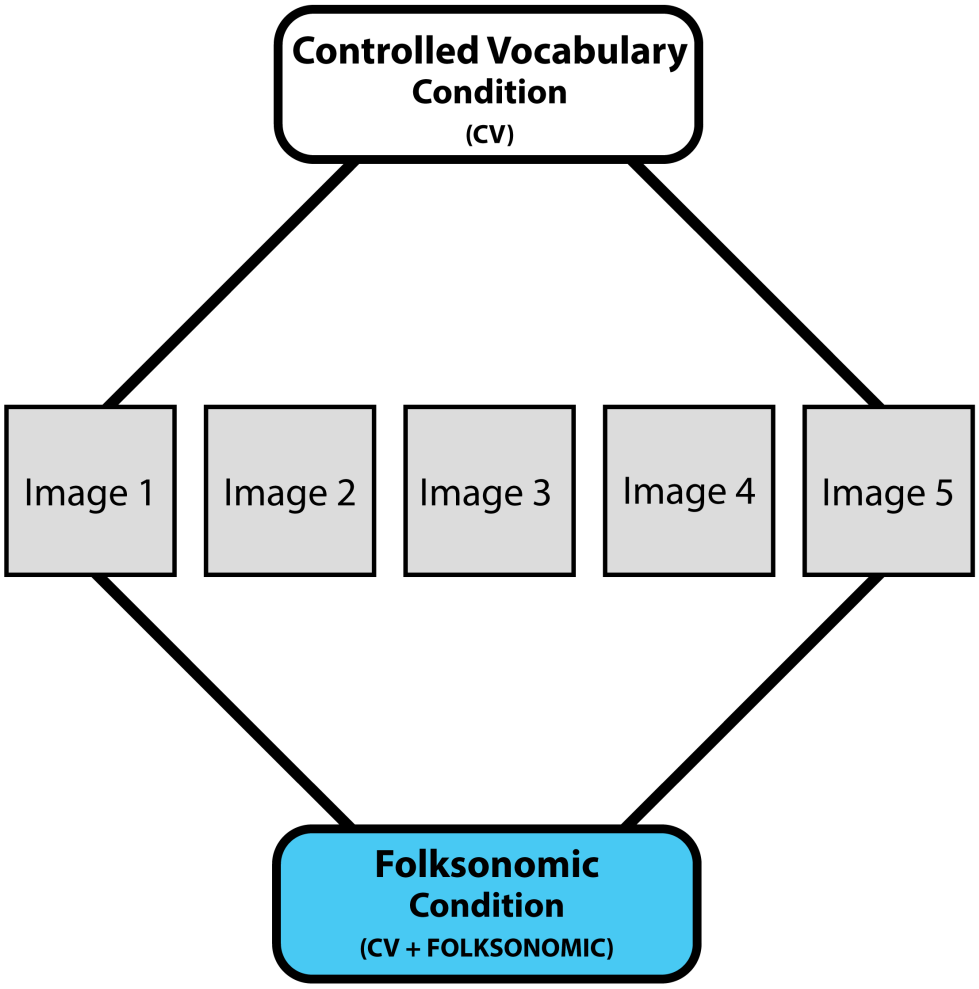

Figure 6. Schematic representation of study design for Task 1 (Image Search).

Additionally, the system returned exact matches only, which meant that if a participant searched for “sailboat” and the only term present in the system was “sailboats,” the search would be unsuccessful. This aspect of the study design was necessitated by the technical specifications and functionality of the version of the Metadata Games search index used in the study, rather than a strategic methodological choice. The frequency of preventable “exact match” retrieval failures is discussed below in Section 4.1. 13

To further illustrate the differences between the two search index conditions, consider the case of a participant in each condition attempting to retrieve image 3 (see Figure 3 above). In the controlled vocabulary condition, the only search terms that would yield a successful retrieval were: “harbors,” “sailboats,” “marblehead harbor” and “glass negatives.” In contrast, in the folksonomy condition, a participant would successfully retrieve this item by entering any of the following search terms: “harbor,” “sailboats,” “water,” “woman,” “sail boats,” “porch,” “scenic,” “view,” “sail,” “sun,” “summer,” “marblehead harbor,” “boats,” “dock,” “harbors,” “veranda,” “balcony,” “girl looking at boats,” “marina,” “glass negatives,” “sailboats on water” and “yacht.” 14

Immediately following the image search task, participants were instructed to play three different single-player tagging games from the Metadata Games suite: *Zen Tag*, *NexTag*, and *Stupid Robot*. In the “free-tagging” game *Zen Tag* (Figure 7), users are able to input as many tags as they wish for four separate images. *NexTag* (Figure 8), uses the same game mechanic as *Zen Tag*, but utilizes a more minimalist user interface and presents a more robust image to players. Finally, in *Stupid Robot* (Figure 9), a novice robot asks users to help it learn new words by tagging images. The game presents one image to users and gives them two minutes to input tags, with the constraint that they may only enter one tag for each given word length (i.e., one four-letter word, one five-letter word, and so on). In playing a single 15

session of each game, participants in the present study tagged the same nine images (four images each in *Zen Tag* and *NexTag* and one in *Stupid Robot*). Tags from these three games were compiled and compared against the traditional metadata provided by staff from the Boston Public Library.

Figure 7. Zen Tag.

Figure 8. NexTag.

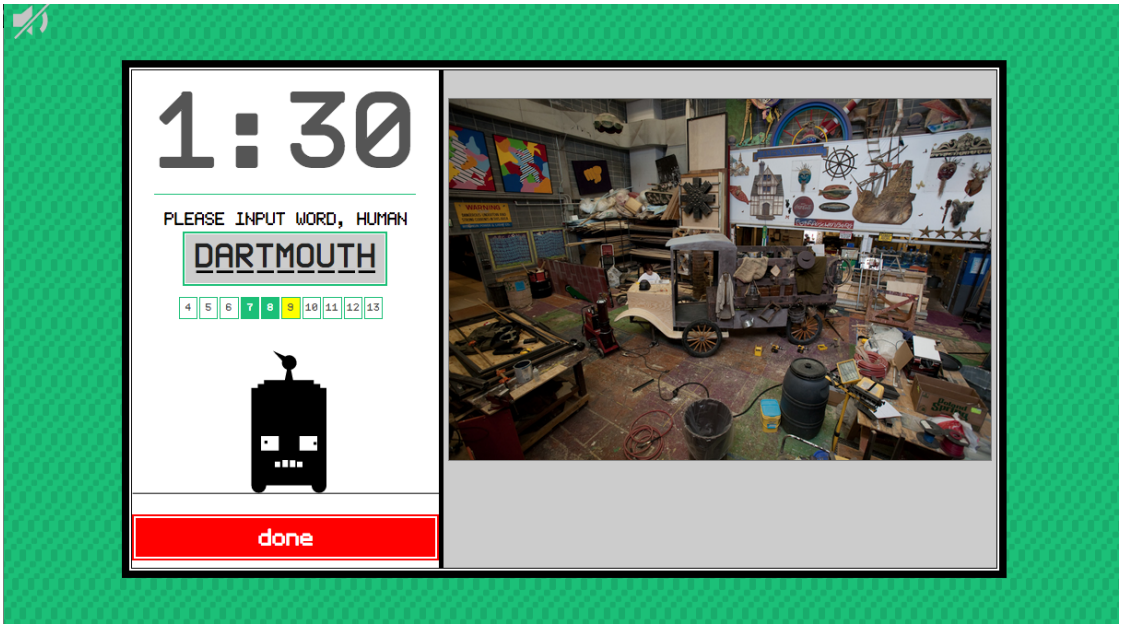

Figure 9. Stupid Robot.

3.1 Scoring

Tags were scored by the lead author using a revised version of the Voorbij and Kipp scale used by Thomas et al. [Thomas et al. 2009]. This scale, which was chosen due to its overall similarity to the Library of Congress Subject Headings hierarchy, includes the following categories for scoring the level of correspondence between a folksonomic tag and a tag included in the controlled vocabulary for the same item: 16

1. Exact match to controlled vocabulary
2. Synonyms
3. Broader terms
4. Narrower terms
5. Related terms
6. Terms with an undefined relationship
7. Terms that were not related at all

A score of one was thus reserved for an exact match between a user-provided tag and the professional metadata, including punctuation. To illustrate, consider the sample image provided in Figure 10 and the corresponding professional and folksonomic metadata provided in Tables 1 and 2 below. With this example, the user-provided term “Hindenburg Airship” would not be deemed an exact match because, as indicated in Table 1, the controlled vocabulary term encloses “Airship” in parentheses. Scores three through five were based on judgments made by the Library of Congress in their subject heading hierarchy. For example, “dog” would represent a broader term of the controlled vocabulary term “Golden retriever,” whereas the tag “biology” would represent a narrower term than the controlled vocabulary term “Science.” We reserved “related terms” (a score of 5) for tags referring to objects or concepts that were represented in the image but not expressed in the professional metadata. A score of six was only to be awarded if, after research, the conclusion was made that the term was unrelated to the image or any of the terms included in the controlled vocabulary. A score of seven, though rare, was reserved for useless “junk” tags, such as any term that was profane, explicit, nonsensical or anything that would not qualify as useful to libraries (e.g., anything under the “Self Reference” or “Task Organization” classes mentioned previously). 17

Figure 10. Hindenburg explodes. (Leslie Jones Collection; Boston Public Library)

Airships
Explosions
Aircraft Accidents
Hindenburg (Airship)
1934-1956 (approx.)

Table 1. Professional Metadata for Figure 9.

Folksonomic Metadata	Score (Voorbij and Kipp Scale)	Notes
Hindenburg (Airship)	1	exact match to “Hindenburg (Airship)”
Hindenburg	2	synonym for “Hindenburg (Airship)”
Accidents	3	broader term of “Aircraft accidents”
Zeppelin	4	narrower term of “Airships”
Flames	5	Present in photograph; related to “Explosions”
Painting	6	this is a photograph
omgreadlater	7	junk tag

Table 2. Folksonomic Metadata and Scores from Voorbij and Kipp Scale for Figure 9.

4. Findings and Analysis

4.1 Findability and Searching

Search Time. On average, participants in the controlled vocabulary index condition took six times longer to search for each image (*M* = 154.1 secs, *SD* = 98.84) compared to participants in the combined index condition (*M* = 25.08 secs, *SD* = 19.39) (see Figure 11 below). A one-way Analysis of Variance (ANOVA) confirmed that this difference was a statistically significant one, *F* (3, 15) = 3.94, *p* < .04. However, it is important to note that participants were free to ask to move on at their discretion if they no longer wished to continue searching for an item. Focusing exclusively on the search times associated with *successfully* retrieved items in both conditions, an ANOVA confirmed that participants in the controlled vocabulary condition still exhibited a significantly longer search time per item (M = 111.36, SD = 89.24) compared to participants in the folksonomy condition (M = 19.94, SD = 9.04), F (1, 12) = 8.51, *p* < .014. 18

Items Found. Because participants were allowed to "give up" on finding any particular item, each participant was assigned a numeric score from 0 to 5 to represent the number of items they *successfully* found. A one-way ANOVA showed that the average number of "found items" was significantly higher in the folksonomy index condition (*M* = 4.88, *SD* = .35) than in the controlled vocabulary index condition (*M* = 1.38, *SD* = 1.06), *F* (1, 15) = 78.40, *p* < .001 (see Figure 12 below). 19

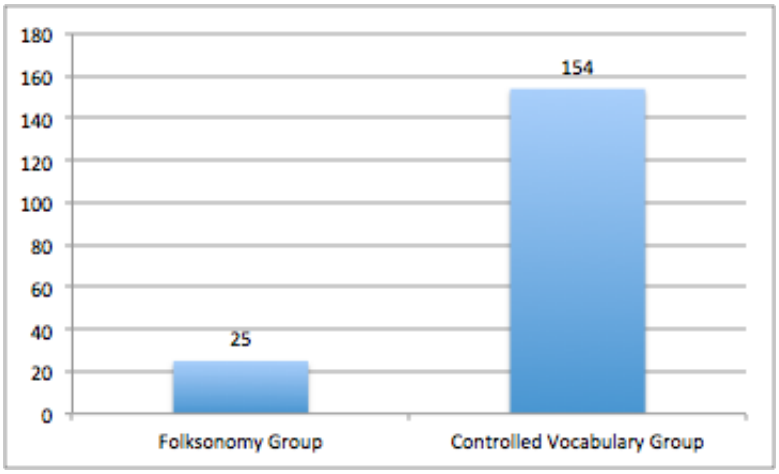

Figure 11. Average Group Search Time (in secs).

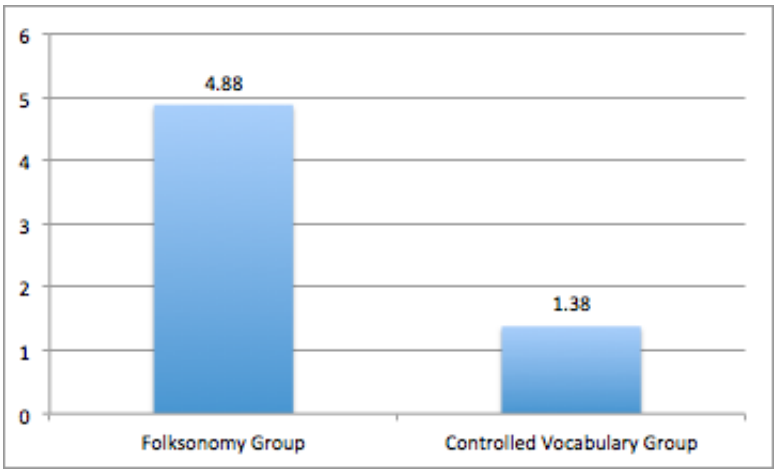

Figure 12. Average Items Found per group.

When taking the failed searches of the controlled vocabulary condition into account, there were a total of 131 completely "preventable" failures overall. Importantly, a majority of these failures were due to the entry of folksonomic synonyms for tags that were included in the controlled vocabulary index, which clearly demonstrates the value of folksonomies for improving search effectiveness and efficiency. Additionally, of the total 379 searches that participants in the controlled vocabulary condition conducted, only a small fraction of search failures (13 searches or 3.4%) were caused by the exact 20

match parameters set forth by the Metadata Games search system. Thus, the differential levels of search time and item retrieval exhibited by participants in the two search index conditions are primarily attributable to the specific tags that were accessible in the two indexes – and not to the particular of the search system used in the present study.

4.2 Tag Analysis

Recall that all tags generated by participants in the gameplay portion of the study were coded using the Voorbij and Kipp scales; Figure 13 (below) depicts the breakdown of scores assigned to the 811 tags generated by the participants. 21

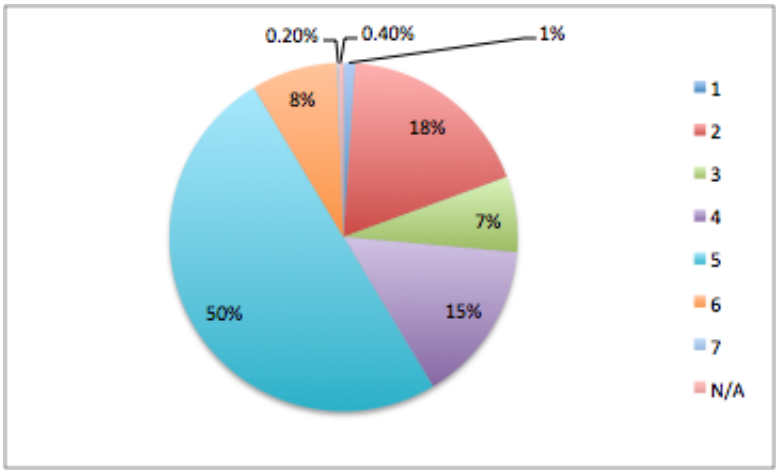

Figure 13. Breakdown of overall Voorbij and Kipp scores.

As shown above, a score of five (“related terms”) accounts for the largest segment of recorded tags, meaning that 50% of all of the tags entered were valid classifications not included in traditional metadata. This implies a fundamental semantic gap between traditional classification and folksonomies. 22

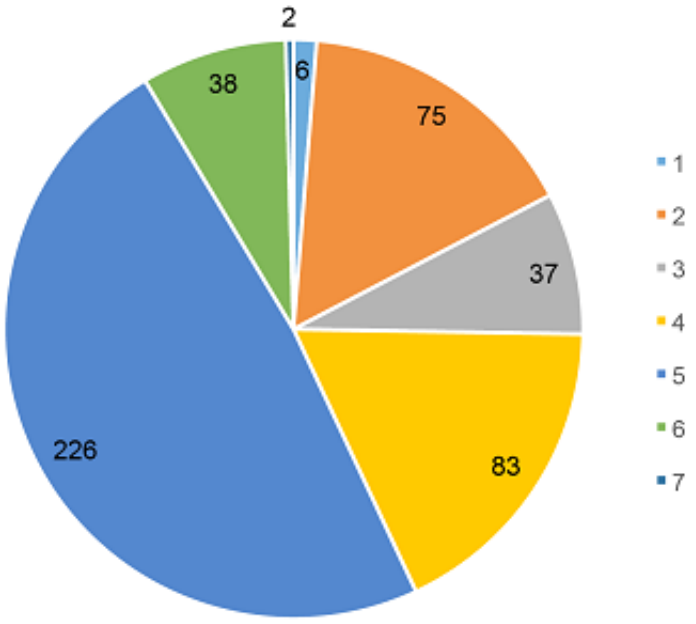

Figure 14. Distribution of Voorbij and Kipp scores for tags generated by LIS participants.

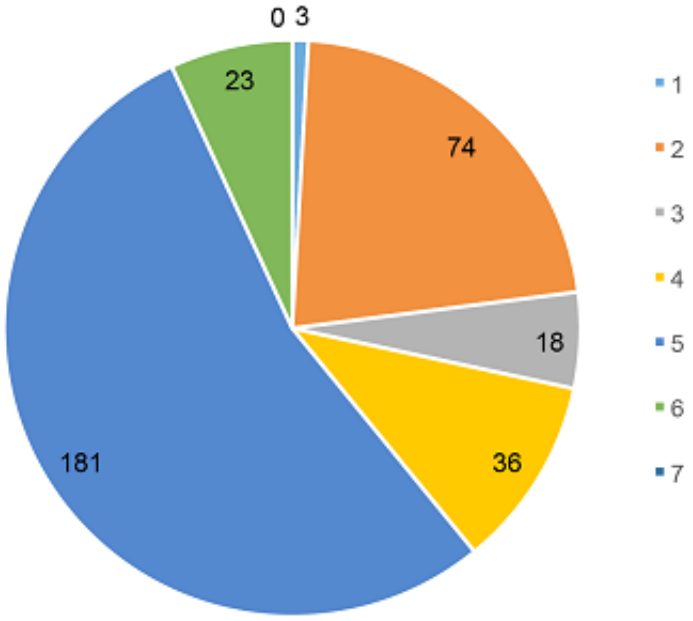

Figure 15. Distribution of Voorbij and Kip scores for tags generated by non-LIS participants.

As illustrated in Figures 14 and 15 (above), the distribution of Voorbij and Kipp scores was constant and nearly identical between the LIS and non-LIS subsamples. This suggests that, when given the same instructions, both librarians and non-librarians can and do produce the same types of useful, accurate data. Additionally, a score of seven, for a so-called “junk” tag, was equally rare in both subsamples’ data. By comparing every participant’s percentage of exact matches and synonyms versus undefined and unrelated terms (Figure 16 below), it is clear that most participants (88%) show an inclination towards folksonomic correctness. It is worth noting that the two highest scoring participants were a horticulturalist between the ages of 18-25 (participant 16) and a librarian between the ages of 45-60 (participant 14). 23

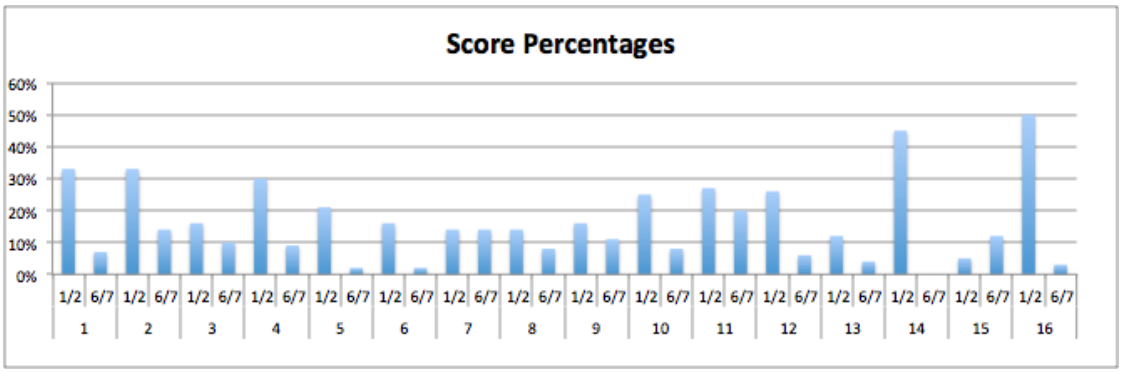

Figure 16. Score percentages for perfect matches and synonyms vs. undefined or unrelated terms.

4.3 Best Subjects for Crowdsourcing

Another concern for cultural heritage institutions is determining what media subjects might work best to collect new metadata through crowdsourcing. As previously mentioned, the images that participants tagged in the present study were divided into five key subject groups (see Figure 17). 24

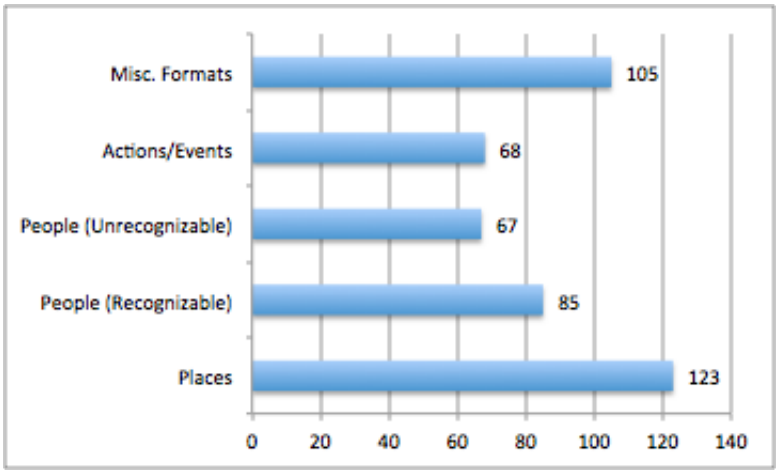

Figure 17. Average Search Time per Image Type (in seconds)

Results revealed that the images garnering the highest number of unique tags were those that fell into the categories People (Unrecognizable) and Miscellaneous Formats (in this case, a digitized newspaper). The image that generated the fewest tags was Image 4, a picture of Thomas Edison, Harvey Firestone, and Henry Ford. Few people recognized the inventors and many simply input tags such as “old men,” although it is important to note that several participants did express some level of familiarity with the figures in the image (e.g., one participant uttered, “I feel like I should know this.”). These results suggest that the best subjects for crowdsourced metadata might be media items that requires no prior knowledge. For example, the unrecognizable person and the digitized newspaper were some of the only instances in which the intent of the photograph was either completely subjective (unrecognizable person) or objectively stated (digitized newspaper). Many other images of famous historical figures and events simply caused the participants to become frustrated with their own lack of knowledge. In light of this fact, crowdsourcing platforms may be well-advised to provide users with the tools to perform their own research about the content of the media to fill in any gaps in knowledge or recollection that they experience. This is a challenge that Metadata Games has begun to address, with the addition of a Wikipedia search bar to encourage users to research what they do not know about a particular media item. 25

5. Conclusions

As of now, there remains debate about the comparative value of traditional and folksonomic metadata as organizational systems for today’s information needs. Nonetheless, there is growing recognition of the fact that folksonomies offer libraries with an ideal return-on-investment scenario [Syn and Spring 2009] with minimum cost (much of which can be off-set by digital humanities grants), maximum output of data [Bischoff et al. 2009] [Noll and Meinel 2007], as well as the chance to increase community engagement with their patrons. As the findings of the present study demonstrate, folksonomic metadata, when used in tandem with traditional metadata, increases findability, corrects preventable search failures, and is by and large accurate. Furthermore, the data suggest that given the same tagging conditions, librarians and non-librarians produce a surprisingly similar distribution of useful metadata. Collectively, these findings point to the potential to change the way we search for and organize our most treasured media. 26

Acknowledgements

The research team would like to thank all study participants, Tom Blake, Linda Gallant, the Boston Public Library, The Digital Public Library of America, Mary Wilkins Jordan, Jeremy Guillette, and the National Endowment for the Humanities. 27

Notes

[1] Other noteworthy examples include the steve.museum, one of the first projects to crowdsource metadata in the cultural heritage domain, which created an online web interface for registered users to submit tags to a selection of works from participating museums

[https://code.google.com/p/steve-museum/]. In addition, in 2008, the Library of Congress (LoC) used the photo-sharing service Flickr to gather comments and tags on 3,000 images [Springer et al. 2008]. From this project, the LoC teamed with Flickr to form the Flickr Commons, a communal page for other cultural heritage institutions with image collections [https://www.flickr.com/commons].

Works Cited

21st Century Lexicon folksonomy. (n.d.). Dictionary.com's 21st Century Lexicon. Retrieved March 31, 2015, from Dictionary.com website: http://dictionary.reference.com/browse/folksonomy

American Library Association - Allied Professional Association 2008 American Library Association - Allied Professional Association (2008). Salary Survey 2008. Retrieved April 11, 2014 from http://www.ala.org/educationcareers/employment/salaries.

Antell and Huang 2008 Antell, K., and Huang, J. (2008). Subject Searching Success Transaction Logs, Patron Perceptions, and Implications for Library Instruction. *Reference and User Services Quarterly 48, no. 1 (Fall 2008):* 68-76. Academic Search Complete, EBSCOhost

Bischoff et al. 2009 Bischoff, K., Firan, C.S., Paiu, R., Nejdl, W., Laurier, C. and Sordo, M. (2009). Music Mood and Theme Classification-a Hybrid Approach. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009),* pages 657-662.

Flanagan et al. 2013 Flanagan M., Punjasthitkul, S., Seidman, M., and Kaufman, G. (2013). Citizen Archivists at Play: Game Design for Gathering Metadata for Cultural Heritage Institutions. *Proceedings of DiGRA 2013, Atlanta, Georgia*. Retrieved January 15, 2014 from http://www.tiltfactor.org/wp-content/uploads2/tiltfactor_citizenArchivistsAtPlay_digra2013.pdf

Golder and Huberman 2006 Golder, S., and Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32(2),* 198–208.

Grey and Hurko 2012 Grey, A., and Hurko, C. (2012). So You Think You're an Expert: Keyword Searching vs. Controlled Subject Headings. Codex: *The Journal of the Louisiana Chapter of the ACRL, 1(4).* Retrieved March 1, 2014 from http://journal.acrlla.org/index.php/codex/article/view/47

Gross and Taylor 2005 Gross, T. and Taylor, A. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries, May 2005.* Retrieved May 1, 2014 from http://crl.acrl.org/content/66/3/212.full.pdf+html.

Hedden 2008 Hedden, H. (2008). How Semantic Tagging Increases Findability. *Econtent Magazine*. Retrieved March 1, 2014 from http://www.hedden-information.com/SemanticTagging.pdf

Heintz and Taraborelli 2010 Heintz, C., and Taraborelli, D. (2010). Editorial: Folk Epistemology. The Cognitive Bases of Epistemic Evaluation. *Review of Philosophy and Psychology, 1(4)*. doi:10.1007/s13164-010-0046-8

Heo-Lian Goh et al. 2011 Heo-Lian Goh, D., Ang, R., Sian Lee, C., and Y.K. Chua, A. (2011). Fight or unite: Investigating game genres for image tagging. *Journal of the American Society for Library Science and Technology, 62(7),* 1311–1324. doi:10.1002/asi.21478

Kipp 2006 Kipp, M. E. I. (2006). Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator and Intermediary Keywords. *Canadian Journal of Information and Library Science, 29(4),* 419–36.

Lu et al. 2010 Lu, C., Park, J., Hu, X., and Song, I.-Y. (2010). Metadata Effectiveness: A Comparison between User-Created Social Tags and Author-Provided Metadata (pp. 1–10). *IEEE*. doi:10.1109/HICSS.2010.273

Maggio et al. 2009 Maggio, L., Breshnahan, M., Flynn, D., Harzbecker, J., Blanchard, M., and Ginn, D. (2009). A case study: using social tagging to engage students in learning Medical Subject Headings. *Journal of the Medical Library Association, 97(2),* 77.

Mai 2011 Mai, J.-E. (2011). Folksonomies and the New Order: Authority in the Digital Disorder. *Knowledge Organization, 38(2).* Retrieved from http://jenserikmai.info/Papers/2011_folksonomies.pdf

Moltedo et al. 2012 Moltedo, C., Astudillo, H., and Mendoza, M. (2012). Tagging tagged images: on the impact of existing annotations on image tagging (p. 3). *ACM Press.* doi:10.1145/2390803.2390807

Nielsen 2012a Nielsen, J. (2012). Usability 101: Introduction to Usability. *Nielsen Norman Group*. Retrieved December 15, 2014 from http://www.nngroup.com/articles/usability-101-introduction-to-usability/.

Nielsen 2012b Nielsen, J. (2012). How many test users in a usability study? *Nielsen Norman Group.* Retrieved April 30,

2014 from http://www.nngroup.com/articles/how-many-test-users/.

Noll and Meinel 2007 M. G. Noll, C. Meinel (2007). Authors vs. Readers: A Comparative Study of Document Metadata and Content in the World Wide Web. *Proceedings of 7th International ACM Symposium on Document Engineering, Winnipeg, Canada, August 2007,* pp. 177-186, ISBN 978-1-59593-776-6

Oomen 2011 Oomen, J., and Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. In *Proceedings of the 5th International Conference on Communities and Technologies* (pp. 138–149). New York, NY, USA: ACM. doi:10.1145/2103354.2103373

PEW Internet 2013 PEW Internet (2013). Library Services in the Digital Age Part 2: What people do at libraries and library websites. PEW Internet. Retrieved January 11, 2015 from http://libraries.pewinternet.org/2013/01/22/part-2-what-people-do-at-libraries-and-library-websites/.

Peters 2009 Peters, I. (2009). Folksonomies: indexing and retrieval in Web 2.0. Berlin: De Gruyter/Saur. Retrieved March 1, 2014 from http://books.google.com/books?isbn=3598441851

Porter 2011 Porter, J. (2011). Folksonomies in the library: their impact on user experience, and their implications for the work of librarians. *The Australian Library Journal.* Retrieved March 1, 2014 from http://www.tandfonline.com/toc/ualj20/current#.UxTDvPRdVLo.

Springer et al. 2008 Springer et al. (2008). For the Common Good: The Library of Congress Flickr Pilot Project Washington. *Library of Congress*. Retrieved on January 11, 2015 from http://www.loc.gov/rr/print/flickr_report_final.pdf.

Stone 2000 Stone, A. (2000). The LCSH Century: A brief history of the Library of Congress Subject Headings, and introduction to centennial essays. *Cataloging & Classification Quarterly 29(1-2).* Retrieved December 23, 2014 from http://www.cataloging andclassificationquarterly.com/ccq29nr1-2ed.htm.

Syn and Spring 2009 Syn, S.Y. and Spring, M.B. (2009). Tags as Keywords – Comparison of the Relative Quality of Tags and Keywords, In *Proceedings of ASIS&T Annual Meeting, 46(1).* November 6-11, 2009, Vancouver, BC, Canada.

Syn and Spring 2013 Syn, S.Y., and Spring, M.B. (2013). Finding subject terms for classificatory metadata from user-generated social tags. *Journal of the American Society for Library Science and Technology, 64(5),* 964–980. doi:10.1002/asi.22804

Thomas et al. 2009 Thomas, M., Caudle, D. M., and Schmitz, C. M. (2009). To tag or not to tag? *Library Hi Tech, 27(3),* 411–434. doi:10.1108/07378830910988540

U.S. Constitution U.S. Constitution, pmbl.

Voorbij 1998 Voorbij, H. J. (1998). Title keywords and subject descriptors: a comparison of subject search entries of books in the humanities and social sciences. *Journal of Documentation, 54(4),* 466–476. doi:10.1108/EUM0000000007178