In the Margins:
Reflections on Scribbles, Knowledge Organization, and Access

June Abbas

Department of Library and Information Studies,
State University of New York at Buffalo, 534 Baldy Hall, Amherst,
NY 14260-1020, <abbasjm@buffalo.edu>

Dr. June Abbas is an Associate Professor in the Department of Library and Information Studies at the State University of New York at Buffalo. Her teaching and research interests allow her to pursue her passion for increasing our understanding of the interconnected facets of access to accurate, relevant information including: 1) how people use information systems and their processes for finding information, 2) how information systems are organized and the schema used to organize the information or representations of information objects that are included in the system, 3) user-centered design, and 4) the changing nature of information and systems. She has also been the project manager for six digital library projects developed by her students for small cultural heritage organizations.

ABSTRACT: Marginalia or 'scribbling in the margins" is a means for readers to add a more in-depth level of granularity and subject representation to digital documents such as those present in social sharing environments like *Flickr* and *del.icio.us*. Social classification and social sharing sites development of user-defined descriptors or tags is discussed in the context of knowledge organization. With this position paper I present a rationale for the use of the resulting folksonomies and tag clouds being developed in these social sharing communities as a rich source of information about our users and their natural organization processes. The knowledge organization community needs to critically examine our understandings of these emerging classificatory schema and determine how best to adapt, augment, revitalize existing knowledge organization structures.

1. Scribbles in the margins

A favored text, dog-eared and yellowed from use, yet still useful, brings back insights that we try to impart to our students when we teach knowledge organization, or organization and control of recorded information courses, whichever words we have chosen to label them. Scribbled in the margins are remnants of notes to ourselves – keywords, subject headings? "tags"? – to remind us of why a particular passage was relevant to us. (In no manner do I condone the practice of writing in the margins of library books; rather, I present this as an analogy.) These scribbles include notes about the thoughts, subjects, eloquent linguistics that we wish to remember, and to access at a later time, maybe even our thoughts that oc-

curred as we read the words. These interactions with the text extend our own personal narrative and create new meaning for us. Should someone pick up this same text and read the passages and also the notes, would one necessarily draw the same conclusions, or would one have yet other insights into the author's meanings, the scribbles, the words? Wilson (1968, 18) reminds us that "what a text says is not necessarily what it reveals or what it allows us to conclude ... but what is not said may interest us more than what is said."

Marginalia, or the practice of writing notes in the margin, has provided us with insight into readers' thought processes, historical context at the time of reading, and greater levels of granularity as the reader engages with the authors of the text. I am reminded

of the impact on scientific discovery accomplished by reading someone else's notes in the margins. Johannes Kepler's work on elliptical orbits was influenced by notes he read in the margins of a second-hand copy of Copernicus' *De revolutionibis* (Gingerich 2004). John Adams, a significant figure in United States and world history, was well known for creating extensive marginalia in his books. His most highly annotated book is Mary Wollstonecraft's *An historical and moral view of the origin and progress of the French Revolution*, which contains 10,085 words of Adams's handwritten commentary. According to the Boston Public Library's (2006) online collection of his works, his marginalia provides us with "a permanent record of John Adams's intensive and frequent interactions with his library. Adams read with a pen clasped firmly in hand and many of his books – particularly those he read late in life – are filled with passionate commentary and lively dialogues with the authors in the margins."

How then do we access the facts, truths, or assertions that the text conveys, or doesn't convey, or the different truths or assertions that occur to another who reads the text? Our knowledge organization structures provide access points to follow. Classification schemes, controlled vocabularies, ontologies, taxonomies, and the like, have been used to access various levels of subject content within the texts. (The use of the word "texts" for this discussion could include any information-bearing object, regardless of format, but to maintain the "argument" being developed, the word "text" will be used throughout.) How then, do we access the "meaning," the conclusions, or the insights others make while reading the words, the scribbles in the margins?

2. Knowledge organization is not static

This is an old argument. Knowledge organization structures are not static. We struggle to update classification schemes. We conduct research to determine if they work. Controlled vocabularies have been criticized as being out of date, containing arcane, discriminatory, Anglo-centric terminology (Olson 2002). We have conducted studies that show that users don't understand how to use subject headings (Markey 1984; Drabenstott and Vizine-Goetz 1990), or that the words they choose for searching do not match subject headings (Taylor 1984; Carlyle 1989; Doyen and Wheeler 1989; Lester 1989; Abbas 2002). So what have we done with the knowledge we

gained from this research? Has it changed our way of thinking about knowledge organization and subject access? How do we apply these understandings to the digital environment?

On the surface it seems the Web has taken much of knowledge organization out of our hands. Users can access this vast depository of texts by entering a few words into a search box, and they do. Studies have shown us that most web searchers are not concerned with thinking up precise, well defined Boolean search strings. They enter a few key (relevant to them) words and click a button. They then sift through the multitude of hits and find at least one or more that satisfices their information need. In online collaborative sharing communities, such as *Flickr* (http://www.flickr.com), *del.icio.us* (http://www.delicious.com), and *LibraryThing* (http://www.librarything.com), users can organize images, cluster bookmarks, and catalog their own personal libraries using words that are relevant to them. The Pew Research Center has estimated that 28 percent of online Americans have tagged content on the Internet, and 7 percent say they tag or categorize content on at least a daily basis (Rainie 2007). These "taggers" are not using our knowledge organization devices. They are creating their own as they use and view others' tags. Vander Wal (2006) has been credited with coining the term "folksonomy," for the resulting cluster of terms that emerges when a community describes texts. Folksonomies are then used for subject representation by the users within these collaborative sharing communities. Again, this idea is not new to knowledge organization, but it is one that has not been widely accepted or applied. Earlier, "pre-tagging" proponents of the process of enabling subject access using user-defined descriptors are: Hastings (1995); O'Connor (1996); Bates (1998); O'Connor, O'Connor and Abbas (1999); Abbas (2002, 2005a, 2005b) to name a few.

3. Tagging as Annotation

The Web has, however, given us an environment to test the efficacy of using user-defined descriptors for subject (as well as physical) access. We might then assume that collaborative sharing communities are in effect, "scribbling in the margins" when users tag their images, their bookmarks, their libraries. This may seem like a weak analogy to some. However, what we do know about the reasons and uses of annotation in the print environment can also be extended to the digital tagging practices as well. For

74

Knowl. Org. 34(2007)No.2
J. Abbas. In the Margins: Reflections on Scribbles, Knowledge Organization, and Access

example, Ovsiannikov, Arbib & McNeil (1999) found that the primary uses for annotation of print were to think, to remember, to clarify, and to share. A further study by Fu, Ciszek, Marchionini, & Solomon (2005) of web-based annotation tools found that the uses or motivations for annotation were similar in the web environment to those in the print environment. While the Fu et al. study did not look specifically at social sharing communities' tagging practice as annotation, the web tools they did examine (personalized web spaces like *MyLibrary*, or blogs and wikis) have enough features in common to extend their findings to the "tagging as annotation" analogy presented here.

Also note that use of the term "scribbling" in either context (print or digital) should in no way indicate a quick, easy process void of thought or consideration. Some tags may be created quickly, but others are only applied after much deliberation, examination of existing tags, or even by using the tag clouds, or other social classification structures of the community. "Taggers" also report that they have referred to outside sources (eg., *Wikipedia*, or existing controlled vocabularies available online) for terms with which to tag. This anecdotal evidence is is based on conversations I have had with taggers, as well as presentations and discussion sessions given by students for a course-related *Flickr* assignment. Users are choosing a few words or phrases to represent the "meaning" of the text to them. They are then re-using these words as their own "controlled vocabularies." While these personalized controlled vocabularies are not following any set of pre-established conventions or standards with which we are comfortable, nor are they either hierarchical in nature or able to show relationships with which we are familiar, (i.e. broader, narrower, related terms), the users have chosen these terms as those found useful for describing the texts. While their process of selecting terms may vary, though at this point we know very little about their selection and application processes, it cannot be assumed that no forethought goes into their selections.

Further, it is also common practice in social collaborative communities to invite other members to provide their own additional tags to your texts, thereby providing yet another dimension of meaning, another level of granularity, to the representation. Tag clouds (the resulting visual structures built as a result of tagging objects) can then become visual representations of meaning to at least one user, the micro-communities or interest groups that form around an interest in the subject, and to a larger so-ciety of users. Tag clouds, in effect, become mechanisms not just for representation, but for retrieval as the individual returns to the tag clouds as a means for retrieving objects, or finding others' related objects that have been tagged with the same terms. Visual representation of search terms in conjunction with index terms, as well as the user's ability to view or manipulate proximity relationships between the two, is not new to knowledge organization and information retrieval. Korfhage (1991) and others (Kim & Korfhage 1994; Nuchprayoon & Korfhage 1994) began exploring visual representation in their work on Visual Information Retrieval Interfaces (VIRI's) and others have followed. VIRI's allow the user to view their search terms and the index terms of the document collections and to set the parameters of how closely the two sets must match for precise retrieval. Tag clouds, while not exactly the same process, are produced by the user, or community of users, and are then the source for sorting, searching, and browsing a collection.

Blair (1990; 2006) provides a further context for examining social representation and access issues. He posits that the language we use both to represent our information needs and to index texts is learned in a social context or community. Blair explains the theory of "language games," as first developed by the early twenty century philosopher Ludwig Wittengenstein, and the process in which we learn language and meaning. We do not acquire language purely by learning the word and its definition, but instead learn its use and appropriateness within the context of our "forms of life" or everyday experiences. Furthermore, we have to possess some prior understanding of the form of life or the language game context we are engaged in before the words can have meaning. Users of online sharing communities are engaging within the social context of a particular community. Each person who contributes tags is engaging in "language games" as they go through their daily "forms of life" or experiences. Where this practice may differ from Wittgenstein's conception, is that there are few limits on what is accepted or unaccepted practice in the social sharing communities. In fact this lack of standards or constraints is highly prized by users who do not want to spend time learning rules. Users can tag using their own constructions, experiences, meanings, with the only limits imposed being of a technological nature.

This is not to say that clusters of meaning or community practices (norms) are not evolving and emerging from this seemingly chaotic environment.

Much of the tagging literature focuses on the resulting folksonomies that are being developed. Much of that same literature is quick to point out the problems of tagging and folksonomy development and use (i.e. tags are sloppy, imprecise, redundant, nonsensical, ambiguous, for single use, no synonym or homonym control, contain both plurals and singular forms, etc.) (e.g.,. Bates 2006, Gordon-Murnane 2006, Golder and Huberman 2006, Guy and Tonkin 2006, Peterson 2006, and MacGregor and McCulloch 2006) and is hesitant to discuss the potentials it can afford us as engineers of knowledge organization, educators, scholars, and researchers. We are beginning to discuss the practical applications or pedagogical necessities of examining this emerging phenomenon (e.g.. Abbas & Graham 2006, Arch 2007, Dye 2006 and 2007, Fichter 2006, Fox 2006, Matusiak 2006, Peterson 2006, Skiba 2006, Spiteri 2006a and 2006b, and Suster 2006).

4. Glimpses of User Perception

So, where does this leave us? Where do we go from here? We have a rich source that is untapped. Our OPACs gather users' search terms and search sessions. Websites also track and collect this same information about access. Online collaborative sharing sites are developing folksonomies, "forming" tag clouds. Each of these sources can tell us volumes about how our users access information. These sources provide us with a glimpse into user's perceptions, cognitive processes, and strategies as they search for texts or when they tag their objects by "scribbling in the margins". At the very least, these sources provide us with the *terms used*, and with further study, may potentially provide contextual meaning. What we need to consider now is how we can use these sources to adapt, augment, and revitalize our knowledge organization structures. There are efforts underway to do just this. Museum and library communities, for example, are exploring the usefulness, as well as logistics, of gathering and incorporating users' tags into their websites, online exhibits, and WebPACs (Trant 2006, Sweda 2006). Others such as Spiteri (2006a, 2006b) are comparing the tags developed by users in *del.ci.ous*, *Furl*, and *Technorati* to existing knowledge organization structures such as the *Library of Congress Subject Headings*. Digital libraries that have been developed for youth are also exploring the idea of using user-defined descriptors as subject headings (Abbas 2002, 2005a, 2005b, and Reuter & Druin 2004).

More needs to be considered. More needs to be learned. We need to step back and critically assess the current state of knowledge organization and its efficiency in the digital environment. We also must ask "What do we really know about social classification, tagging, and its meaning and use for users? How can social classification and tagging practices inform knowledge organization practices and structures?" Some potential areas to explore include:

What does tagging mean to users? Is tagging a way to describe a text, a scribble in the margins, or does it provide a set of search terms? Are these potential uses (description, search terms) different to users?

What are users' motivations for tagging (personal findability or organization; communal or familial sharing; meaning making; performative act)? Do motivations affect user's choices of tags?

Can we apply Wittgenstein's "Language Games" theory to what is happening in online sharing communities? And if so, how can this lens inform knowledge organization theory and practice?

What can we learn from collaborative classification, folksonomy development? How can we incorporate this learning into classification scheme and controlled vocabulary development? Should we try to make tags more consistent and require users to follow knowledge organization conventions or do we just watch and learn? Can or should we apply traditional controlled vocabulary constraints to user-defined descriptors?

What are the implications to knowledge organization and library and information science education? How do we incorporate social classification tools, practices, and user's changing expectations and motivations into our curricula?

The growing body of literature on this topic suggests that "tagging" and these digital "scribbles in the margins" are not a fad or trend that will soon pass. This phenomenon is providing us with a much needed glimpse into our users' natural organization processes. Social representation and sharing sites are affording us a rich, complex dataset to work with, to examine more contextually, to serve as a means for us to critically analyze our existing schema and from

76

Knowl. Org. 34(2007)No.2
J. Abbas. In the Margins: Reflections on Scribbles, Knowledge Organization, and Access

which to learn, to adapt, to generate new meaning to our existing knowledge organization structures. If we don't take advantage of this opportunity, we might find others scribbling in the margins about our lack of involvement, and how we were left behind and became obsolete.

References

Abbas, June. 2002. Smoothing the information seeking path: Removing representational obstacles in the middle school digital library environment. Ph.D. dissertation, University of North Texas, Denton.

Abbas, June. 2005a. Out of the mouths of middle school children: I. developing user-defined controlled vocabularies for subject access in a digital library. *Journal of the American Society for Information Science and Technology* 56: 1512-24.

Abbas, June. 2005b. Creating metadata for children's resources: issues, research, and current developments. *Library trends* 54(2): 303-17.

Abbas, June and Graham, Jennifer E. 2006. So, let's talk about tagging, user-defined/supplied descriptors …. a research and curricular agenda. In *Information realities, shaping the digital future for all: Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology, Austin, TX, November3-8, 2006*. Medford, N.J.: Information Today.

Arch, Xan. 2007. Creating the academic library folksonomy: put social tagging to work at your institution. *College & research libraries news* 68(2): 80-81.

Bates, Marcia J. 1998. Indexing and access for digital libraries and the Internet: human, database, and domain factors. *Journal of the American Society for Information Science* 49: 1185-1205.

Bates, Mary Ellen. 2006. Tag – you're it! *Online* 30(1): 64.

Blair, David. 1990. *Language and representation in information retrieval*. Amsterdam: Elsevier Science.

Blair, David. 2006. *Wittgenstein, language, and information: back to the rough ground!* Dordrecht: Springer.

Boston Public Library. 2006. *The John Adams Library at the Boston Public Library*. (http://www.johnadamslibrary.org/, accessed June 27, 2007).

Carlyle, Alyson. 1989. Matching *LCSH* and user vocabulary in the library catalog. *Cataloging & classification quarterly* 10(1/2): 37-63.

Doyen, Sally E. and Wheeler, Daniel D. 1989. Use of a controlled vocabulary index in information re-

trieval tasks. In Salvendy, Gaviel and Smith, Michael J., eds. *Proceedings of the Third International Conference on Human-Computer Interaction on Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, 2nd ed. New York: Elsevier Science, pp. 226-31.

Drabenstott, Karen and Vizine-Goetz, Diane. 1990. Search trees for subject searching in online catalogs. *Library hi tech* 8(3): 7-20.

Dye, Jessica. 2006. Folksonomy: a game of high-tech (and high-stakes) tag. *EContent* 29(3): 38-43.

Dye, Jessica. 2007. Collaboration 2.0. *EContent* 30(1): 32-36.

Fichter, Darlene. 2006. Intranet applications for tagging and folksonomies. *Online* 30(3): 43-45.

Fox, Robert. 2006. Cataloging for the masses. *OCLC Systems & Services* 22: 166-72.

Fu, Xin, Ciszek, Tom, Marchionini, Gary and Solomon, Paul. 2005. Annotating the Web: an exploratory study of Web users' needs for personal annotation tools. In *Sparking synergies: bringing research and practice together; proceedings of the 68th Annual Meeting of the American Society for Information Science & Technology, October 28-November 2, Charlotte, NC*. Medford, N.J.: Information Today.

Gingerich, Owen. 2004. *The book nobody read: chasing the revolutions of Nicolaus Copernicus.* NewYork: Penguin Books.

Golder, Scott and Huberman, Bernardo. 2006. Usage patterns of collaborative tagging systems *Journal of information science* 32: 198–208.

Gordon-Murnane, Laura. 2006. Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher* 14(6): 26-38.

Guy, Marieke and Tonkin, Emma. 2006. Folksonomies: tidying up tags? *DLib magazine* 12(1). Retrieved June 27, 2007 from http://www.dlib.org/dlib/january06/guy/01guy.html.

Hastings, Samantha K. 1995. Query Categories in a Study of Intellectual Access to Digitized Art Images. In Kinney, Tom, ed. *Converging technologies, forging partnerships in information, proceedings of the 58th ASIS Annual Meeting: Chicago, October 9-12, 1995*. Medford, NJ: Information Today, pp.3-8.

Kim, Hanhwe and Korfhage, Robert. 1994. BIRD: Browsing interface for the retrieval of documents. In *Proceedings of the IEEE Symposium on Visual Languages*. Los Alamitos: IEEE Computer Society Press, pp. 176-177.

Korfhage, Robert. 1991. To see, or not to see – is that the query? In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research*

and Development in Information Retrieval. New York: ACM Press, pp. 134-41.

Lester, Marilyn Ann. 1989. Coincidence of user vocabulary and *Library of Congress Subject Headings*: experiments to improve subject access. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Macgregor, George and McCulloch, Emma. 2006. Collaborative tagging as a knowledge organisation and resource. *Library review* 55: 291-300.

Markey, Karen. 1984. *Subject searching in library catalogs: before and after the introduction of online catalogs*. Dublin, OH: Online Computer Library Center.

Matusiak, Krystyna. 2006. Towards user-centered indexing in digital imaging collections. *OCLC Systems & Services* 22. 283-98.

Nuchprayoon, Assadaporn and Korfhage, Robert. 1994. GUIDO, a visual tool for retrieving documents. In *Proceedings IEEE Symposium on Visual Languages 1994*, pp. 64-71.

O'Connor, Brian C. 1996. *Explorations in indexing and abstracting: pointing, virtue and power*. Englewood, CO: Libraries Unlimited, Inc.

O'Connor, Brian., O'Connor, Mary K. and Abbas, June. 1999. User reactions as access mechanism: an exploration based on captions for images. *Journal of the American Society for Information Science* 50: 681-97.

Olson, Hope. 2002. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic.

Ovsiannikov, Ilia. A., Arbib, Michael A. and McNeil, Thomas H. 1999. Annotation technology. *International journal of human-computer studies* 50: 329-62.

Peterson, Elaine. 2006. Beneath the metadata. *D-Lib magazine* 12(11).

Rainie, Lee. 2007. 28% of online Americans have used the internet to tag content: forget Dewey and his decimals, internet users are revolutionizing the way we classify information—and make sense of it. *Pew Internet and American Life Project*. Retrieved June 27, 2007, from http://www.pewinternet.org/pdfs/PIP_Tagging.pdf.

Reuter, Kara and Druin, Alison. 2004. Bringing together children and books: an initial descriptive study of children's book searching and selection behavior in a digital library. In *Managing and enhancing information: cultures and conflicts, proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*. Medford, NJ: Information Today, pp. 339-48.

Skiba, Diane J. 2006. Web 2.0: next great thing or just marketing hype? *Nursing education perspectives* 27: 212-14.

Spiteri, Louise F. 2006a. The use of folksonomies in public library catalogs. *The serials librarian* 51: 75-89.

Spiteri, Louise. 2006b. The use of collaborative tagging in public library catalogues. Available online: http://www.oclc.org/research/grants/presentations/2006/spiteri.ppt.

Suster, Mark. 2006. Folksonomy. *AIIM e-doc magazine* 20(6): 20-21.

Sweda, Jennifer E. 2006. Using social bookmarks in an academic setting: PennTags. Poster presented at the 17th Annual ASIS&T SIG/CR Classification Research Workshop Saturday, November 4, 2006, Austin, TX.

Taylor, Arlene G. 1984. Authority files in online catalogs: An investigation of their value. *Cataloging & classification quarterly* 4(3): 1-17.

Trant, Jennifer. 2006. Social classification in art museums: steve.museum. Panel paper presented at the 17th Annual ASIS&T SIG/CR Classification Research Workshop Saturday, November 4, 2006, Austin, TX.

Vander Wal, Thomas. 2006. Folksonomy to improve IA. Presented at the OZ IA, Sydney,Australia. Retrieved May 14, 2007, from http://s3.amazonaws.com/2006presentations/OZIA/Folksonomy_for_IA.pdf.

Wilson, Patrick. [1968] 1978. *Two kinds of power: an essay in bibliographical control*. California library reprint series ed. Berkeley: University of California Press.