Exploring the potential for social tagging and folksonomy in art museums: proof of concept

JENNIFER TRANT

Archives & Museum Informatics / Faculty of Information Studies, University of Toronto, Canada

with the participants in the steve.museum project

http://www.steve.museum

Correspondence: J. Trant, Email: jtrant@archimuse.com

Documentation of art museum collections has been traditionally written by and for art historians. To make art museum collections broadly accessible, and to enable art museums to engage their communities, means of access need to reflect the perspectives of other groups and communities. Social Tagging (the collective assignment of keywords to resources) and its resulting Folksonomy (the assemblage of concepts expressed in such a cooperatively developed system of classification) offer ways for art museums to engage with their communities and to understand what users of on-line museum collections see as important. Proof of Concept studies at The Metropolitan Museum of Art compared terms assigned by trained cataloguers and untrained cataloguers to existing museum documentation, and explored the potential for social tagging to improve access to museum collections. These preliminary studies, the results of which are reported here, have shown the potential of social tagging and folksonomy to open museum collections to new, more personal meanings. Untrained cataloguers identified content elements not described in formal museum documentation. Results from these tests – the first in the domain – provided validation for exploring social tagging and folksonomy as an access strategy within The Metropolitan Museum, motivation to proceed with a broader inter-institutional collaboration, and input into the development of a multi-institutional collaboration exploring tagging in art museums. Tags assigned by users might help bridge the semantic gap between the professional discourse of the curator and the popular language of the museum visitor. The steve collaboration (http://www.steve.museum) is building on these early studies to develop shared tools and research methods that enable social tagging of art museum collections and explore the utility of folksonomy for providing enhanced access to collections.

Keywords
social tagging, folksonomy, steve, museum, social engagement, visitor experience, user experience

1. Providing Access to Museum Collections

Our collective cultural and natural heritage is embodied in the collections of museums. These institutions assemble, group and present artefacts in ways that communicate this meaning, recognizing that information about objects is often as important as the artefacts themselves (Boas 1907). Museums have been experimenting with and using technology to communicate knowledge about their collections over the past five decades (The Metropolitan Museum of Art 1968; Ellin 1969). This development of digital museum programs, reflected for example, in the papers of the International Cultural Heritage Informatics Meetings (Bearman 1991; through Perrot 2005) and Museums and the Web conferences (Bearman and Trant 1997; through Trant and Bearman 2006) has taken place in the context of an increasing openness and an awareness of museums' diverse roles in a heterogeneous community that is shaped by the philosophies of the "new museology" (Anderson 2004).

Museums were drawn to the Web by the ease and economy of publishing collections documentation (including large numbers of color images). But models for organizing and presenting complex cultural heritage information on-line are still being explored. The portions of museum Web sites that focus on collections tend to be either highly-authored, linear exhibitions and educational materials such as lesson plans, or un-interpreted collections databases. Authored materials tend to have a very strong museum "voice" and institutional perspective (Walsh 1997). In contrast, collections databases present the characteristics of objects (such as creator, size, materials, use, provenance,) without context, and in isolation from related works. This isolation is problematic because, typically, museum collections are comprised of very large numbers of objects that to the "un-trained eye" can seem very similar. Fossils, chairs, textiles, works of modern art: it takes the knowledge and perspective of a specialist to distinguish one thing from the next. When many similar things are briefly described in a database, what results is an homogeneity that is difficult to penetrate. The knowledge organization systems that underlie museum documentation reflect specialist perspectives and museum business processes (for example McKenna and Patsatzi 2005; Canadian Heritage Information Network (CHIN) [2005]) not public perceptions and interests.

Neither the authored narrative nor the database of collections information fully supports museums' goals of improving access and building understanding of the objects in their care. Collections are available, but not really accessible. They may be described, but aren't readily comprehensible. Things that might seem exceptional to the general viewer – that a painting is *of a cow looking at a painting* – might not be mentioned at all in traditional museum documentation:

Department	Modern Art
Title	The Innocent Eye Test
Classification	Paintings
Artist	Mark Tansey, American, born 1949
Date Label	1981
Medium	Oil on canvas
Dimensions	78 x 120 in. (198.1 x 304.8 cm)
Credit Line	Partial and Promised Gift of Jan Cowles and Charles Cowles, in honor of William S. Lieberman, 1988
Accession Number	1988.183
Source	*The Metropolitan Museum of Art, Collections Documentation Database Record (1988.183)*

One way museums have begun to facilitate a more personal encounter with collections has been through on-line personal galleries. For example, both the Fine Arts Museums of San Francisco (Fine Arts Museums of San Francisco 2004-) and The Metropolitan Museum of Art, New York (The Metropolitan Museum of Art 2005 -) allow visitors to create personal collections that assemble user-selected works from the museum and share them as on-line exhibitions. The philosophical issues about this kind of program, including the challenge posed to the museum's responsibility to interpret its collections (and to its sense of authority) have been reviewed in the context of the release of myVirtualGallery at the Art Gallery of New South Wales, Australia; interestingly, the final iteration of this on-line gallery does not include the word 'curator' (Cooper 2006).

Within this context, *social tagging* – where "tags" (user-supplied access points for works of art, or other information resources supplied and shared by the general public (museums' audience) in a common on-line environment – and *folksonomy* – the resulting "socially constructed classification" system – are appealing to museums as they appear to fill gaps in current documentation practice and enable a desired level of social engagement (Mathes 2004; Smith 2004; Hammond, Hannay et al. 2005; Quintarelli 2005). Tagging enables a departure from the authored voice of the museum and, through the distributed contribution of many individuals, the construction of additional means of access to collections. For art museums, acknowledging such alternative perspectives is a significant departure that reflects a growing understanding of museums' place in a diverse community.

2. Museums, Tagging and Community

Social tagging is one of a number of Internet-based technologies museums have used to encourage public engagement with their collections (Durbin 2004). Users are enabled to use personal narrative to define the significance of museum artifacts in projects such as Every Object Has a Story (Victoria & Albert Museum, Ultralab et al. 2005 -). These projects derive their user-centred approach from constructivist educational theory that emphasizes individual meaning-making as central to personally significant encounters on-line and in-gallery (Hein 1998; Samis 1999). Social Tagging appeals to museums partly because it embodies this philosophy: tagging represents a dialog between the viewer and the work, and the viewer and the museum (Bearman and Trant 2005). Taken individually, a tag is a user's assertion that a work of art is *about* something (in some way, at some time). In the museum context, tagging offers a way for people to connect directly with works of art, to own them by labeling or naming them – one of the aspects of sensemaking (Golder and Huberman 2006). Tagging lets users assert their own connections and associations between objects in ways that reflect personal perspectives and interests. Tagging further enables re-discovery of works previously seen; users' tags record salient characteristics of personal interest and support subsequent searches (Chun, Cherry et al. 2006).

Tagging in a museum context may differ from other implementations of social tagging (including the shared bookmarking services such as del.icio.us [http://del.icio.us] and Connotea [http://www.connotea.org/]) because museums have existing relationships with visitors that define a social and cultural context for the tagging activity. Museums invest in programs for teachers and students, and volunteers and docents; tagging could become part of the museum's tool-set for fostering and maintaining these relationships. Tagging could also facilitate teacher or student use of the collection; volunteers could actively tag as part of their contribution to the institution. Rather than being motivated by personal gain (Vander Wal 2005), these kinds of users donate time and knowledge. The Cleveland Museum of Art plays on such an altruistic motivation in the links to its on-line tagging tool that say "Help others find this object" (Cleveland Museum of Art and Hiwiller 2005). At the Powerhouse Museum, in Sydney Australia, the Electronic Swatchbook project (Powerhouse Museum and Chan 2005) appears to violate one of the roles of social tagging – to provide immediate feedback – by collecting terms for museum review and future use.

Perhaps the most significant challenge tagging poses to the museum is that it is a visitor-initiated activity; the viewer of a work supplies its significance. Tagging represents an investment in the museum's collection by an individual. The visitor adds value for themselves, for the museum, and other visitors by revealing different perspectives and contexts. These enhance, and possibly subvert, institutional perspective. Folksonomies constructed in social tagging environments are direct evidence of what people see as significant. Looking at the types of tags supplied by those outside museums and studying how they correlate (or don't) with data now made available by museums can provide insight into users' perceptions, identify areas of disconnect, and help museums adapt to meet their missions.

3. Bridging a Semantic Gap

Interpreting works of art to the general public requires bridging the semantic gap between the professional, curatorial language of art history and the public perceptions of its visual evidence. This is one of the great challenges in teaching Art History 101. The gap is reflected in the way that searches are made of public art resources (McCorry and Morrison 1995; Roberts 2001; Sundt 2002; Choi and Rasmussen 2003). Including folksonomic terms in museum records would seem a promising way to provide access along facets of interest to the general public.

But professionals in art museums initially met such a suggestion with skepticism. As collection description and interpretation is a prime professional role of the museum, ceding it to the general public seemed an abrogation of responsibility. Museums see themselves as the source for authoritative information about the objects in their care. Social tagging appears to undermine that role, letting 'just anybody' play a role in how a work is described and ultimately perceived. Presenting the results of social tagging in an unfiltered manner could result in 'wrong' information being linked to museum works, compromising one of the museum's prime responsibilities. The

irony of this position, however, is that most art museums do not incorporate subject description in their collections documentation. Subject access is not a core requirement for museum business functions (such as registration, inventory and location control, exhibition and loan management), so isn't a core requirement in museum documentation guidelines such as the MDA's Spectrum (McKenna and Patsatzi 2005), or the CHIN Humanities Data Dictionary (Canadian Heritage Information Network (CHIN) [2005]). Though subject access is recommended in more user-oriented documents (such as Visual Resources Association (VRA) Data Standards Committee 2002; Visual Resources Association 2005), it is acknowledged as expensive and difficult (Shatford 1984).

4. Proof of Concept Testing at The Metropolitan Museum of Art

To develop an understanding of the role social tagging might play in the art museum, The Metropolitan Museum of Art conducted a series of proof of concept tests between the Fall of 2004 and the Fall of 2005, to determine if untrained cataloguers could provide useful description and access points through tagging-like activities. The tests were performed to validate the proposition that social tagging could add value to existing museum documentation. The audience for the results was narrowly conceived as internal to the museum, and possibly to the art museum community. However, growing interest in the role of social tagging and folksonomy, and the need for foundational studies that compare different methods of enabling access to on-line collections, motivated our reporting of the results here.

It is tempting to compare this proof-of-concept work to studies of naïve image indexing, but to do so may be to read too much into these preliminary results. The tests were not structured to elicit categories of image description terms (as was the case in Jörgensen's study (1999)) or to build an understanding of the searching habits of users looking for images (as have for example Markey 1988; Sormunen, Markkula et al. 1999; and Jörgensen and Jörgensen 2005 among others, in addition to the art-specific studies cited above). The goal was simply to determine if there was a significant difference between the access points in existing collections documentation – such as artist, title, date, medium/support, dimensions – and the terms that are supplied when viewers described the visual elements of an image and what it 'literally' depicts. Results from these tests – the first in the domain – provided validation for exploring social tagging and folksonomy as an access strategy within The Metropolitan Museum, motivation to proceed with a broader inter-institutional collaboration, and input into the development of a multi-institutional collaboration exploring tagging in art museums that became known as steve (http://www.steve.museum).

4.1 Tests 1 and 2 with Paper and Pencil

The first Proof-of-Concept test was conducted at The Metropolitan Museum of Art, New York in the Fall of 2004, by Daniel Starr of the Museum's Watson Library. It "compared cataloguing by trained library cataloguers and library admin[istration] staff" (Chun 2006). A group of participants (5 Library professionals and 5 Library Assistants/Associates) were shown images 5 works from the museum's collection and asked to record terms on worksheets. Aggregated data from Test 1 is available on-line (The Metropolitan Museum of Art 2005a).

The second Proof of Concept test, also led by Daniel Starr of the Watson Library, was conducted December 7, 2004. A group of ten works was shown to Library Assistants/Associates, and they were asked to record terms on worksheets that prompted for entries in a number of categories: topical; person(s) or groups of persons (real or fictional); allegory/symbolism/theme; place; chronology; events; emotions; and other. The terms collected are available on-line (The Metropolitan Museum of Art 2005b).

4.2 Results from Tests 1 and 2

4.2.1 Tagging Provided Additional Access Points

In the first two preliminary tests, tagging provided access points that supplemented and enhanced the technical and curatorial descriptions available in The Metropolitan Museum of Art Collections Database and on the Museum's Web site. (Because museum collection documentation is an on-going process, the level of documentation for each work varied; not all works have full curatorial descriptions, nor do all appear in the Museum's on-line collection or the Timeline of Art History

both linked from http://www.metmuseum.org.) As an example, one of the works included in the December 2004 test was an early 20ᵗʰ century photograph showing Madison Square, New York, from above (Figure 1).

Figure 1. Alvin Langdon Coburn (British, born America, 1882–1966), *The Octopus*, 1912. The Metropolitan Museum, New York (1987.1100.13). Ford Motor Company Collection, Gift of Ford Motor Company and John C. Waddell, 1987

The technical description of Figure 1 reads:

> Alvin Langdon Coburn
> (British, born America, 1882–1966)
> *The Octopus*, 1912
> Platinum print; 41.8 x 31.8 cm (16 7/16 x 12 1/2 in.)
> Ford Motor Company Collection, Gift of Ford Motor Company and John C. Waddell, 1987 (1987.1100.13)

The art historian describes this work's style and technique on The Metropolitan Museum of Art's Web site:

> Couched in the soft velvety nap of the platinum paper, composed in the languid lines of Art Nouveau, and softly focused, this photograph of New York's Madison Square employs many elements of Pictorialism at its best. However, the dizzying effect of Coburn's aerial view and his fascination with the skyscraper are distinctly and precociously modern. The blend of Pictorialist technique and fresh vision was characteristic of the transitional moment when Alfred Stieglitz,

Coburn, Karl Struss, and Paul Strand began to celebrate contemporary urban experience. (The Metropolitan Museum of Art 2006)

But when asked in the December 2004 test to supply terms that described what was depicted in this image, staff from The Metropolitan Museum of Art provided fifty-seven unique terms, listed in Table 1.

1. 20th century	30. outdoors
2. abstract	31. park
3. abstraction	32. park in winter
4. aerial	33. park-goers
5. aerial topography	34. parks
6. areal perspective	35. paths
7. black and white	36. pedestrians
8. black and white contrast	37. photography (b/w)
9. cities	38. public spaces
10. city	39. roads
11. cityscape	40. shadow
12. cityscape in winter	41. shadow (tower)
13. Coldness	42. shadows
14. contrasts	43. sledders
15. empty park	44. Sledding
16. Flat Iron Building	45. sleighs
17. geography as art	46. snow
18. geography in art	47. snowscape
19. landscapes	48. street scene
20. Late 19th/early 20th century	49. street scenes
21. Madison Square	50. tower (shadow)
22. Madison Square (New York)	51. trees
23. New York	52. urban
24. New York City	53. urban landscapes
25. New York City in winter	54. view from a window
26. New York City winter	55. walking
27. New Yorkers	56. Winter
28. NY	57. winter
29. octopus	

Table 1: Terms for *The Octopus* (Figure 1) supplied by staff of The Metropolitan Museum of Art in a proof-of-concept tagging test, December 7, 2004. (The Metropolitan Museum of Art 2005b, Image #9)

Reviewing these terms in their raw form raises all the questions about synonymy, orthography, and controlled vocabulary that characterize discussions of the utility of folksonomy (Shirky 2005; Shirky, Butterfield et al. 2005; Guy and Tonkin 2006). What makes them of particular interest in the context of this test is the different perspective they represent and the number of new concepts found in these terms in comparison to the museum's documentation.

4.2.2 Untrained Cataloguers Provided Useful Terms

The Fall 2004 test compared terms supplied by five Librarians to terms supplied by five Library Assistants and Associates (non-professional library staff). Overall, Library Assistants and Associates provided slightly more terms than the Librarians: for the group of 5 works, Library Assistant and Associates provided a total of 138 terms (or 54.8% of the total number) Librarians 114 terms (or

45.2% of the total). While Librarians assigned an average of 4.6 terms per work, Library Assistants and Associates assigned an average of 5.5 terms per work.

Library Assistants and Associates assigned a larger percentage of the distinct terms as well (see Chart 1). Of the total of 133 distinct terms assigned to the five works, Librarians assigned 39 unique terms, Library Associates 55 unique terms, and both groups 39 terms. While some of these additional terms can be attributed to errors in identification of picture elements by the un-trained Library Assistants and Associates (using cherubim for putti for example, or Nativity for Madonna and Child), a review of the terms suggested by the untrained group showed that they still made significant contributions to improving access (describing both subject elements, such as antelopes, dogs and chains, and emotive qualities such as defiance, despair, and heroic).

There appeared to be slightly more inter-participant consistency within the group of Library

The Metropolitan Museum of Art: Proof of Concept Test 1: Distinct Terms by Participant Type

Librarians
29%

Library Assistants +
Associates
42%
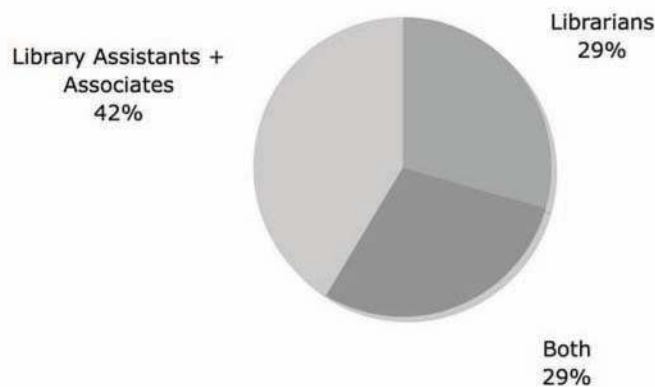
Both
29%

Chart 1: The Metropolitan Museum of Art, Proof of Concept: Test 1:
Distinct terms assigned by Librarians, Library Assistants + Associates, and both groups.

Assistants and Associates, as opposed to the professional librarians. When the number of times a term was assigned to a work is plotted (see Chart 2), a higher number of Library Assistants and Associates assigned the same term four and five times.
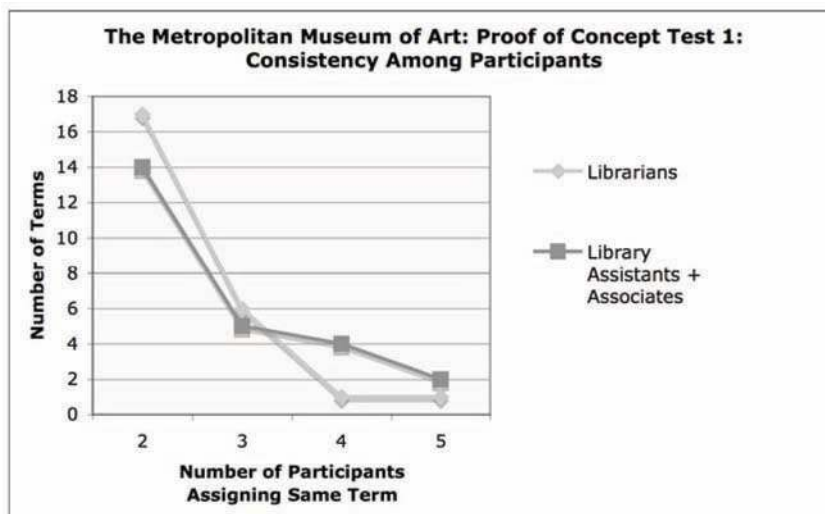

The Metropolitan Museum of Art: Proof of Concept Test 1: Consistency Among Participants

Number of Terms

Librarians

Library Assistants + Associates

Number of Participants Assigning Same Term

Chart 2: The Metropolitan Museum of Art, Proof of Concept Test 1: Consistency between
Participants, comparing Librarians and Library Assistants + Associates.

4.3 Tests 3 and 4 with a Desktop Tool

The positive indications of the potential for social tagging strategies in the first round of tests motivated the development of a more formal Proof of Concept test at The Metropolitan Museum of Art in the summer of 2005. Groups of museum administrative staff and volunteers (with no specific relationship to the museum) were shown a total of 30 images and asked to provide terms that they would use to search for each work of art. These tests were conducted using an automated desk-top term-collection tool developed under the direction of the Museum's Subject Cataloguing Committee by Koven Smith, an analyst in the Information Systems and Technology department (The Metropolitan Museum of Art and Smith 2005). The tool presented a single image of a work of art at a time (see Figures 2-4). Terms were entered in one of 15 input boxes arranged along the left-hand side of the screen under a "Keyword" prompt. Data was recorded in rows that included: Participant Key; Object Key; Caption (text if shown); Keyword (entered); Category (if prompted or none); and Type (whether the image was shown alone, with caption or with category). Tests took place during the summer and fall of 2005; groups of 4-6 participants met in The Metropolitan Museum of Art's Information Technology Training Room, and were supervised by a member of the Subject Cataloging Committee (Jenkins 2006c). Proctors reported some competitiveness between participants that might have been encouraged by this small space. The awareness of how others were doing spurred some to enter more terms and discouraged others, who felt they could not 'keep up'.

The 30 works were divided into three groups, constructed as a preliminary probe into the influence of the tagging environment on the terms supplied. The first group of images was shown without any identifying information (screenshot in Figure 2); the second group was shown with museum 'caption' information, similar to the caption for Figure 1 (screenshot in Figure 3); the third group of images was shown without a caption, but participants were prompted with a category within which to provide terms. The Category was randomly chosen from: Colors; Emotions; Events; People (real or fictional); Places; Themes; Things; and Time (screenshot in Figure 4); no attempt was made to match categories to works as that would require a level of effort that would not scale. Participants were given 15 minutes to provide terms for each group of ten images.

**Figure 2: The Metropolitan Museum of Art Proof of Concept Tests 3 and 4:
Screen Shot from Testing Environment: Image Alone**

Keyword

Submit Keywords   15

Wang Hui (Chinese, 1632–1717)
The Kangxi Emperor's Southern Inspection Tour, Scroll Three: Ji'nan to Mount Tai (Kangxi
nanxun, juan san: Ji'nan zhi Taishan)
China
datable to 1691–98

Figure 3: The Metropolitan Museum of Art Proof of Concept Tests 3 and 4:

Category - People (real or fictional)
Keyword

Submit Keywords   24

Figure 4: The Metropolitan Museum of Art Proof of Concept Tests 3 and 4:
Screen Shot from Testing Environment: Image With Random Category

4.4 Results from Test 4: Fall 2005

Data from tests conducted in the Fall 2005 have been analysed in detail. (Data for the summer 2005 tests is unavailable.)

4.4.1 Terms per Participant

The 39 test participants assigned a total of 6679 terms to the 30 works in the test, an average of 6.5 terms per work. But the number of terms was not distributed equally across all types of presentation. While the same proportion of terms was assigned to works with and without captions, only about half as many terms were assigned to works shown with a random category prompt (Chart 3).
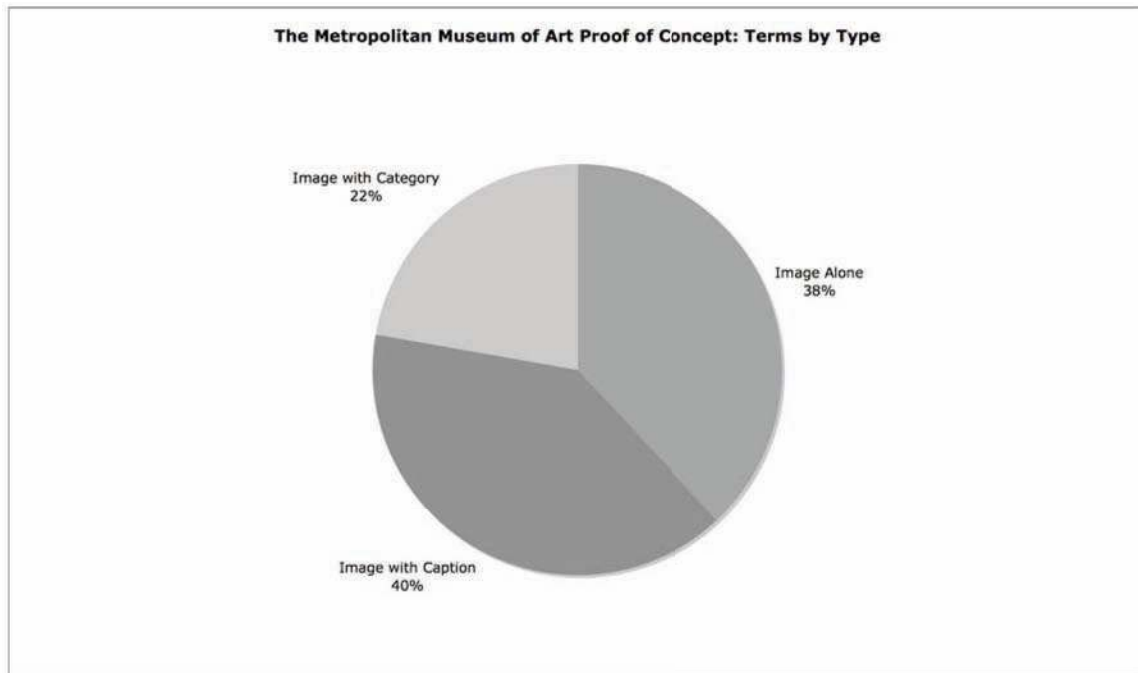
Chart 3: The Metropolitan Museum of Art Proof of Concept: Total Terms Assigned

Metropolitan Museum of Art staff posit that a mis-match between the categories prompted at random and the works shown accounts for some of this discrepancy; this is supported by the fact that taggers were much more likely to skip (i.e. not provide any tags for) works in the Image with Category group. The 35 taggers that completed 80% of the images skipped 4.9% of the images in the group Image with Categories (Table 2). A closer look at the data shows that the averages hide the fact that 17% of the taggers skipped one particular work – a non-representational Arabic manuscript, the *Tughra of Sultan Sulaiman the Magnificent* – when prompted to assign terms in categories such as "Places", "Themes", and "Emotions".

Type	Total	# terms	Avg. per work	# taggers	# works skipped	% works skipped
Image Alone	2482	70.9	7.1	35	10	2.9%
Image with Caption	2656	75.9	7.6	35	4	1.1%
Image with Category	1483	42.4	4.2	35	17	4.9%
Grand Total	6621	189.2	6.3	35	31	8.9%

**Table 2: The Metropolitan Museum of Art Proof of Concept Test: Terms Per Participant,
for those who tagged 80% of the works.**

A look at the number of tags assigned to each group of images by each participant (Chart 4) shows that some participants assigned noticeably more tags to Images with Captions, some assigned noticeably more tags to Images Alone, but all participants assigned fewer tags to Images with Category prompts.
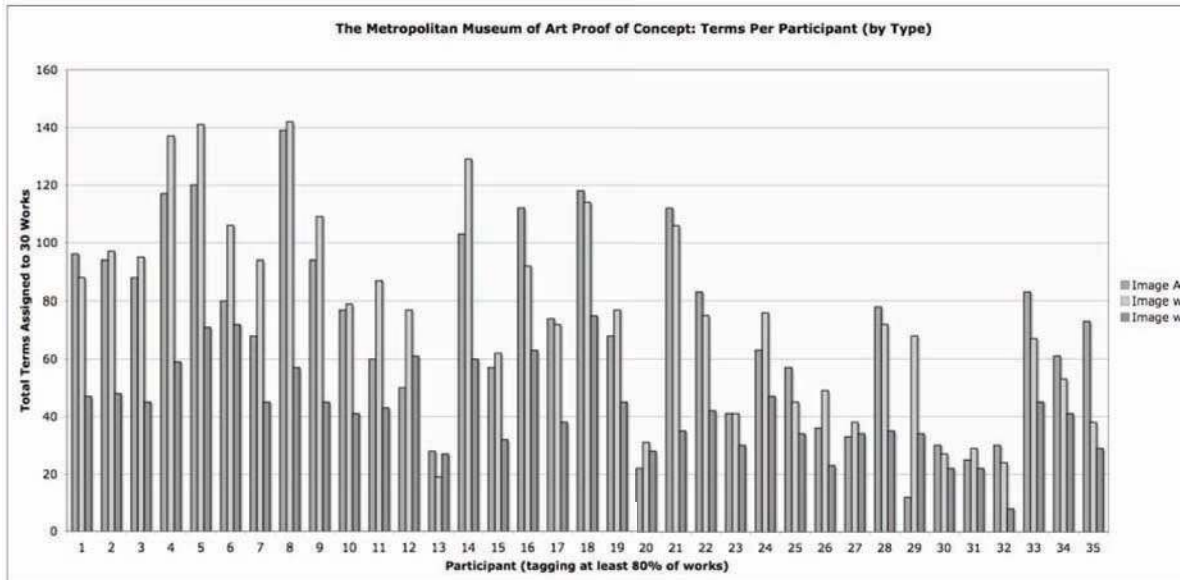

Chart 4: The Metropolitan Museum of Art Proof of Concept: Number of Tags by Participant, broken out by type of image presentation
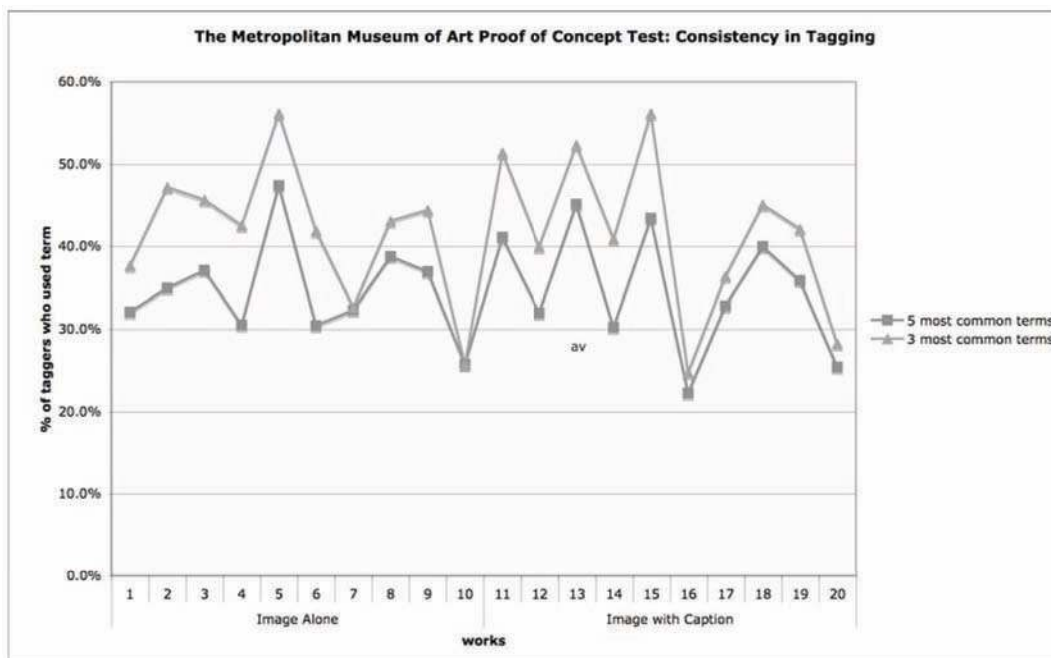
4.4.2 Inter-Participant Consistency

Chart 5: The Metropolitan Museum of Art Proof of Concept: Consistency in Tagging, showing the percentage of cataloguer agreement for the 5 most common, and 3 most common terms.

There was a surprising amount of consistency in the terms that were assigned to works in the groups Image Alone and Image with Caption. (Comparisons of inter-participant consistency are not possible for the Image with Category group because not all participants were prompted for the same category.) The percentage of co-occurrence for the five most common terms (Chart 5) ranged between 22.3% and 47.4%; when the three most common terms are compared, consistency increases

to between 24.8% and 56.2%. When taken across a group of 30 study participants, this is a significant level of agreement; Markey's study reports 2-27% agreement in pre-iconographical descriptors supplied by 39 indexers for 13 different works of art (1984).

Consistency varied across works, indicating that the kind of work that was being tagged could be a factor. Whether or not a caption was present, however, did not influence the consistency of the tags assigned. The average consistency for the five most common terms without captions was 34.7%; with captions it was 34.8%. For the 3 most common terms, the average consistency without captions was 41.7%; with captions it was 41.8 %.

There is some question as to whether consistency between taggers is necessarily desirable. Theoretically, consistency is valued as a quality in cataloguing, as it reflects adherence to professional standards and practices (Shoham and Kedar 2001). The index terms assigned to a document are seen to represent its core content, and consistent indexing is supposed to be key to effective and efficient retrieval (Leininger 2000). But in the context of an art museum actively seeking diversity of perspective, too much consistency might be a bad thing. It may be that, as Bearman argued in the case of archival materials, the difference between the presentation language of the art museum collection user and the language of the documents (or their descriptions) mandates a focus on breadth of description, or enhanced means of access (such as complex lead-in vocabularies) (Bearman 1989). Our goals in tagging are to increase the number of access points, rather than limit indexing to those terms supplied by professionals.
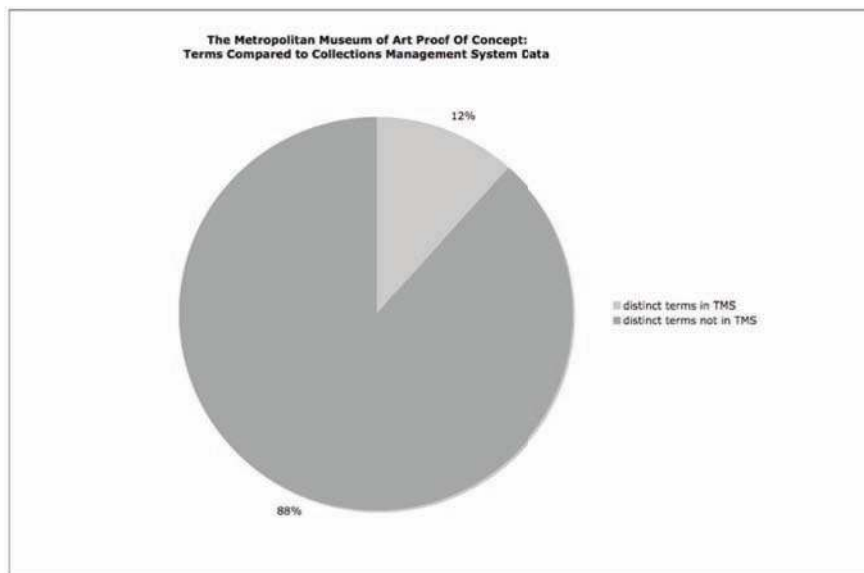

Chart 6: The Metropolitan Museum of Art Proof of Concept: Distinct Terms Compared to Collections Management System Data

One place where tag occurrence analysis may prove fruitful is in the assessment of the relevance of tags to the works described. It may be that further analysis of a larger set of tag data can determine statistically-based thresholds that indicate which tags are appropriate to a work – a concern that can't be discounted in an Internet-based tagging environment open to the possibility of inaccurate or malicious tagging (tag spam). Some review of user-supplied tags is likely to be required in the

museum environment in order to satisfy mission-related concerns for responsibility in collection interpretation.

4.4.3 Evaluation of User-Supplied Terms: Were new terms supplied?

Several analyses were made of the terms gathered in the test to assess the potential for folksonomic terminology collected through social tagging to enhance access to art museum collections. First, to establish if the terms contributed something new to the existing documentation, the terms collected was compared to basic museum documentation (similar to the captions shown above) drawn from the museum's collections management system (fields extracted were Title, Accession number, Object name, Culture, Artist/Maker name, Period, Dynasty, Reign, Style, Medium, Description, Credit line, Department, and Geography [All fields]) (Jenkins 2006b).

Using a simple perl script (authored by Koven Smith) distinct tags per work were matched on a word-by-word basis (ignoring compound terms) to extracted words from the database records. (This simplest comparison would produce the largest number of matches.) There were 3780 distinct terms in the 6679 terms collected. Of these distinct terms, 437 appeared in the Museum's records; 3343 were new. The non-professionals made a contribution: 88% of the terms supplied by participants in the Proof of Concept Tests were not found in the basic museum descriptions (Chart 6).

4.4.4 Evaluation of User-Supplied Terms: Were useful terms supplied?

A critical motivation for the Proof of Concept tests was to learn if non-professionals could provide useful terms that might improve access to art museum collections. The goal was not just more terms (though that improves recall) but terms that accurately described the work of art (ensuring precision). Teams of representatives from The Metropolitan Museum of Art Subject Cataloguing Committee reviewed the tagger-supplied terms that did not appear in collections database records, in order to determine if they were appropriate to the work of art. Appropriate terms that accurately reflected the work of art, were considered 'valid'; terms that reflected mis-readings or inaccuracies were identified as 'invalid'. This can be judged to be a fairly strict assessment of 'relevance' as it was made by museum professionals with art historical training and an interest in art documentation. More than three-quarters of the terms supplied by participants in the proof of concept tests were determined to be valid (Chart 7). This validation was done as part of the Proof of Concept, as an exploration of the differences between user-supplied and professional descriptions, to raise issues that would have to be addressed in the deployment of social tagging of art museum collections, and to build understanding in The Metropolitan Museum of Art's Subject Cataloguing Committee.
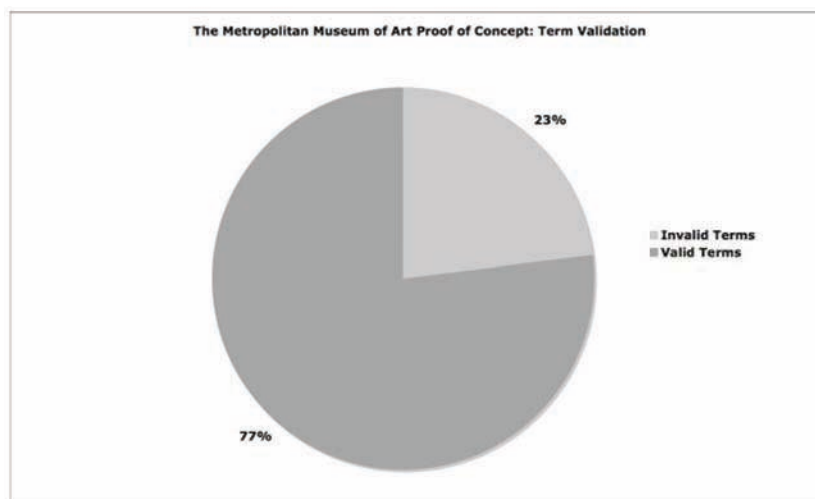

Chart 7: The Metropolitan Museum of Art Proof of Concept: Term Validation Results
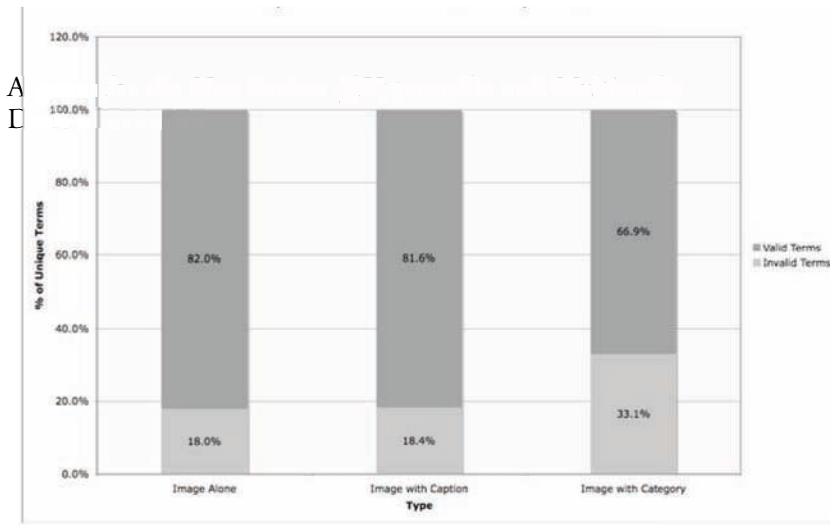
Chart 8: The Metropolitan Museum of Art, Proof of Concept Tests: Term Validation: Percentage of Valid Terms by Test Type.

The proportion of terms judged valid differed depending on whether the image was shown by itself, with a caption, or with a category prompt (Chart 8). About the same percentage of terms were judged as valid if an image was shown with or without a caption (82% vs. 81.6%). However, only 66.9% of the terms supplied in response to images with category prompts were judged as valid. The presence of the category makes comparison across these types difficult. Inconsistencies were also present in the ways that the teams did validation; some validated the term in the category in relation to the work, others – more leniently – validated the term in relation to the work and disregarded whether it was appropriate to the category. (This would raise the number of valid works). Alternate validation methods will need to be explored in future studies.
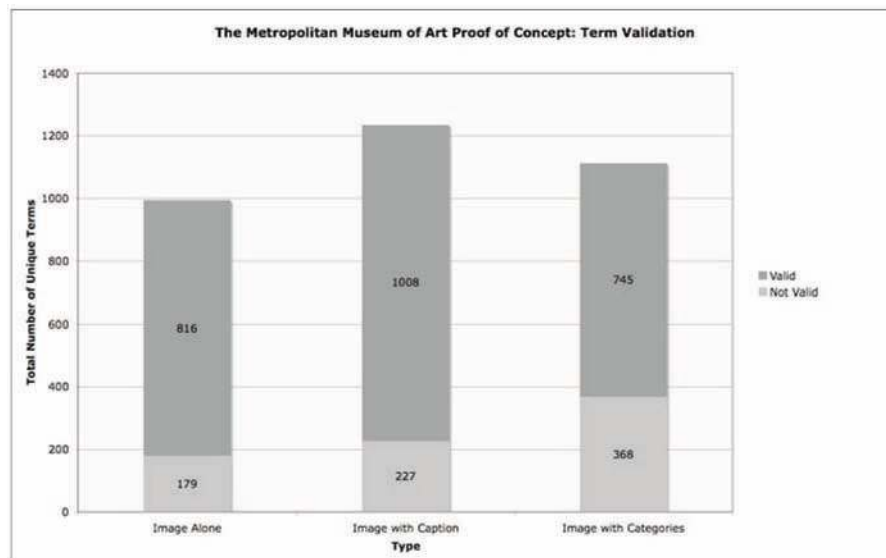

Chart 9: The Metropolitan Museum of Art, Proof of Concept Tests: Term Validation: Number of Terms per Test Type.

When the absolute number of valid and invalid terms are broken out by test type (Chart 9) it becomes clear that more new terms were supplied when an image was shown with its caption. While the overall percentage of terms was about equal in each of these categories (40% vs. 38 %), a larger number of terms to validate (i.e. new terms not found in museum documentation) was produced from the set of works shown with captions: 39% of the terms to validate were assigned to images

with captions, 32% were assigned to images with category prompts, and 29% were assigned to images shown without any associated information (Chart 10).
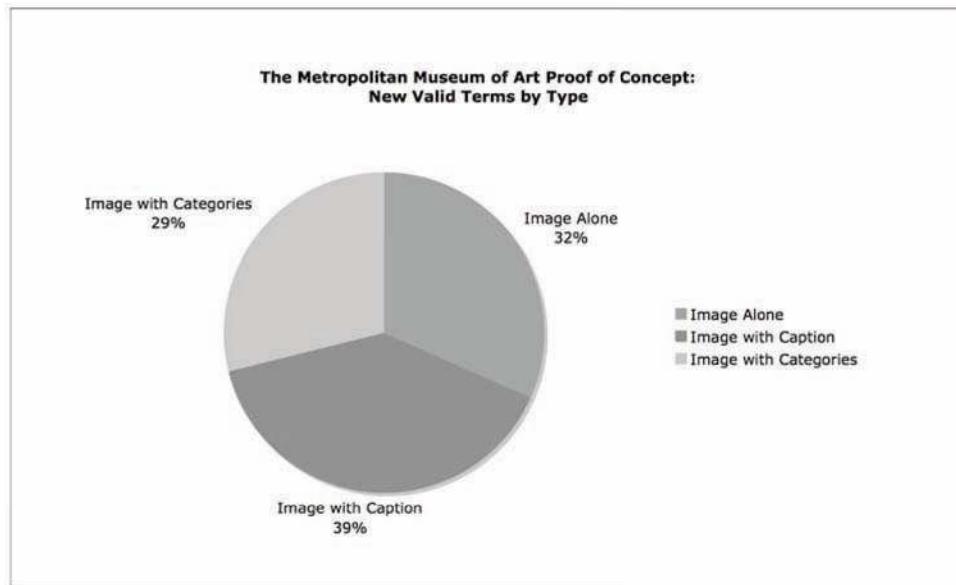

Chart 10: The Metropolitan Museum of Art, Proof of Concept Tests: Term Validation: Percentage of New Terms per Test Type.

4.4.5 Evaluation of User-Supplied Terms: Errors in terms supplied

Some works proved more difficult to describe than others. When the number of valid and invalid terms is compared on a work-by-work basis (Chart 11) it is clear that valid terms out-number invalid terms in almost all cases. Challenging works are easily identified: invalid and valid terms were equal in number in one case (a William Blake drawing that is difficult to read); invalid terms out-numbered valid terms in two cases (the same highly calligraphic Islamic manuscript that many taggers skipped, and an unfamiliar African sculpture with challenging iconography).
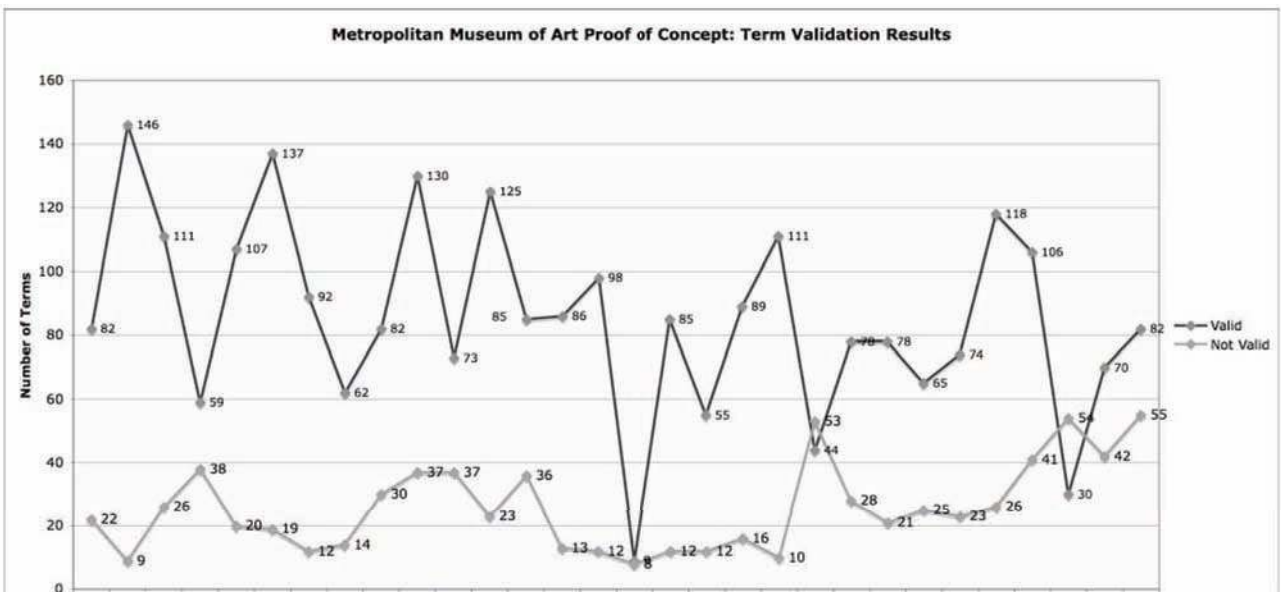

Chart 11: The Metropolitan Museum of Art, Proof of Concept Tests: Term Validation: Numbers of Terms per Work.

These clusters of errors seem to point to 'teachable moments': places where the museum could do more to facilitate the understanding of particular kinds of works. Clusters of invalid tags also indicate where additional interpretation may be required. The most common 'invalid' terms represent consistent mis-readings of particular works: ten participants identified goats as 'sheep'; nine called a sculpture of a Neo-Assyrian mythical beast a 'horse' when it has the form of a winged lion.

5.   Proof of Concept: Conclusions

Museum staff reaction to the terms provided by non-specialist participants in these Proof of Concept tests at The Metropolitan Museum of Art was positive upon analysis. Participants supplied a useful number of terms (7 per work), and of those 77% were judged to be valid. Three times more valid than invalid terms were provided. The tagging results showed high levels of inter-cataloguer consistency: the 5 most common terms were used by an average of 34.8% of taggers; the 3 most common terms were used by an average of 41.8% of taggers. Participant behaviour changed when random categories were introduced into tagging. They supplied fewer terms (even when possible mis-matches between the work and randomly assigned category are taken into account), and many more of the terms assigned to images with a category prompt were judged invalid (even when inconsistencies in validation vis à vis category are taken into account).

Some kinds of works proved harder to tag than others. One work had greater than 60% of the terms assigned to it judged invalid; this was also the most skipped work (the *Tughra of Sultan Sulaiman the Magnificent)* and one of two works that had more invalid than valid terms. In contrast, one work had greater than 96% of the terms assigned to it judged valid. Errors in tagging clustered. The top 3 'invalid' terms represented consistent mis-readings of the work, possibly representing 'teachable moments' or places where museum interpretation is lacking, or reflecting the difficulty of reading digital reproductions of some types of works.

Anecdotal evidence showed some professionals find the basic description of visual elements surprisingly difficult. In an early test at The Metropolitan Museum of Art, a curator exhibited significant discomfort with the standard description task. When asked what was wrong, he blurted out "everything I know isn't in the picture" (Jenkins 2006a). The requirement to put aside previous knowledge may preclude museums from providing some kinds of access points. Museum professionals may not think like naïve users, but they need to reflect these viewpoints in collections documentation. Studies of user queries of art museums show that existing art museum documentation doesn't necessarily answer user questions. A study of user-supplied tags would complement the parsing of queries such as "Do you have any interior scenes by Mr. Rococo that are suitable for framing?" found in the user needs and questions analysis conducted by the Consortium for the Computer Interchange of Museum Information (CIMI) (Janney and Sledge 1995). By looking at tags, museum professionals can see through the eyes of our visitors, respond appropriately, and adapt practices as necessary.

Studies of art museum folksonomies can also supplement knowledge of image retrieval generally. We can build on research in image retrieval (such as (Jörgensen 1999; Chen 2001) with a novel approach that begins with what users notice, declare, and assert as important. The terminological associations derived from tagging would seem just as valid as other techniques (such as card-sorting) for identifying similarity in visual sources. Differences in description might reflect differences in perception that are themselves informative. That users rather than professionals might provide the bulk of indexing for a document could influence cataloguing practice more broadly; it definitely challenges the assumptions underlying most indexing research.

6.   Directions for Future Work: steve.museum

The Proof of Concept exercises at The Metropolitan Museum of Art raised as many issues as they answered, both about folksonomic data collection and analysis methodology, and about the nature and functionality of art tagging tools. A group of professionals from museums and organizations that support them technically and intellectually has formed steve, a collaboration to explore both sides of

the social tagging / folksonomic description question.[1] Coordinated through a shared Research Agenda, (Cataloguing by Crowd Working Group and Trant 2005), the group is developing tools and techniques to support social tagging of art collections, and enable serious study of the resulting folksonomic terminology.

Steve is an open collaboration with an experimental, research-driven methodology. Participants pool raw data, resources and research results in a distributed fashion; each participant addresses those aspects of the shared Research Agenda that mesh with institutional needs, requirements and abilities. The distributed execution of the project reflects diverse (and sometimes almost conflicting) rationales for participation: while some are more interested in the social side of steve, and others in the folksonomy side, the same tools and methods enable the exploration of both. The method is a reflection of the evolving understanding of tagging. The group is consciously exploring options, and has seen its own terminology evolve from a "community cataloguing" or "cataloguing by crowd" focus on the terminology resulting from tagging, to the more inclusive, ambiguous, and non-representational "steve".

7. Conclusion

The Proof of Concept studies at The Metropolitan Museum of Art have supported ongoing development of social tagging tools for use in art museums. They show that non-specialists can supply a useful number of new access points, augmenting the professional descriptions of art museum collections. These preliminary results support investment in further research. Participants in steve have begun crafting experiments that deploy a shared open-source tagging tool and build a common data set to support collective analysis of social tagging and folksonomy in art museums. They do this with a shared goal to encourage user engagement with museum collections, and a shared a mission to create connections between people and art. The museum offers a unique social context within which to explore tagging and folksonomy. Both sides of the dynamic – both the act of naming and the name assigned – provide useful insight into the relationships between museums and visitors. Separately, museums are institutions dedicated to serving local and global communities. Collectively, museums are charged with preserving the cultural trust. Both of these goals require an ability to connect with individuals and communities on their own terms. Works of art are multi-dimensional, but the diverse meanings they communicate are not well understood. Tagging may reveal where museum visitors engage with museum collections. Early, Proof of Concept tests show that tagging is both engaging in itself as it is satisfying to do, and productive as it increases the number of access points to art museum collections.

Museums have invested significantly in the development of standards and systems for documenting museum collections. As with other formal knowledge organization systems, the high overhead of their implementation – both cognitive and institutional – has hindered the broad availability of large amounts of museum data. Social tagging and folksonomy offer a less formal, more participatory, and highly distributed way to augment museums' institutional documentation with content that reflects the perspectives and interests of their communities. Participants in steve are optimistic that they can both engage museum users and learn from them using these new tools.

8. Acknowledgements

---

[1] There is no acronym at the root of steve. See the FAQ at http://www.steve.museum for background on the name.

A paper for the *New Review of Hypermedia and Multimedia*
Draft of 5/13/06

Some of this text draws on a paper prepared for the WWW2006 Workshop on Tagging, May 21, 2006 in Edinburgh, Scotland, co-authored with Bruce Wyman (Trant and Wyman 2006), whose work is reflected in the sections of this paper on the potential for tagging to engage museum audiences.

9. References

Anderson, G., Ed. (2004). *Reinventing the Museum: Historical and Contemporary Perspectives on the Paradigm Shift*. Walnut Creek, CA, Altamira Press.

Bearman, D. (1989). "Authority Control Issues and Prospects." *American Archivist* **52**: 286-299.

Bearman, D., Ed. (1991). *Hypermedia and Interactivity in Museums*. Pittsburgh, Archives & Museum Informatics.

Bearman, D. and J. Trant (1997). *Museums and the Web 97: proceedings* Pittsburgh, Pa., Archives & Museum Informatics. http://www.archimuse.com/mw97/

Bearman, D. and J. Trant (2005). "Social Terminology Enhancement through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections." *D-Lib Magazine* **11**(9). http://www.dlib.org/dlib/september05/bearman/09bearman.html

Boas, F. (1907). "Some Principles of Museum Administration." *Science* **25**(650): 921-933.

Canadian Heritage Information Network (CHIN). ([2005]). "CHIN Data Dictionary: Humanities." Retrieved November 9, 2005, from http://daryl.chin.gc.ca:8000/BASIS/chindd/user/wwwhe/SF.

Cataloguing by Crowd Working Group and J. Trant. (2005, July 23). "Research Questions: A Summary of Issues Raised in Discussion." Retrieved May 2, 2006, from http://www.steve.museum/reference/ComCatResearchQ050723.pdf.

Chen, H.-l. (2001). "An analysis of image retrieval tasks in the field of art history." *Information Processing and Management* **37**(5): 701-720.

Choi, Y. and E. M. Rasmussen (2003). "Searching for Images: the Analysis of Users' Queries for Image Retrieval in American History." *Journal of the American Society for Information Science and Technology* **54**(6): 498-511.

Chun, S. (2006). Email message: "participants in the MMA tagging exercise". to J. Trant, on Sat, 22 Apr 2006 10:04:47 -0400.

Chun, S., R. Cherry, et al. (2006). *Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums*. Museums and the Web 2006: selected papers from an international conference, J. Trant and D. Bearman, Eds. Albuquerque, NM, Archives & Museum Informatics. http://www.archimuse.com/mw2006/papers/wyman/wyman.html

Cleveland Museum of Art and D. Hiwiller. (2005, July). "Cleveland Museum of Art Cataloging by Crowd Prototype Term Collection Tool." Retrieved May 3, 2006, from http://www.steve.museum/reference/ClevelandPrototype-2005.pdf.

Cooper, J. (2006). *Beyond the On-line Museum: Participatory Virtual Exhibitions*. Museums and the Web 2006: Proceedings, J. Trant and D. Bearman, Eds. Albuquerque, New Mexico, Archives & Museum Informatics. http://www.archimuse.com/mw2006/papers/cooper/cooper.html

Durbin, G. (2004, March 25, 2004). "Learning from Amazon and eBay: User-generated Material for Museum Web Sites." *Museums and the Web 2004: Proceedings* Retrieved August 7, 2005, 2005, from http://www.archimuse.com/mw2004/papers/durbin/durbin.html.

Ellin, E. (1969). "Museums and the Computer, an appraisal of new potentials." *Computers and the Humanities* **4**(1): 25-30.

Fine Arts Museums of San Francisco. (2004-, ongoing). "MyGallery." Retrieved January 9, 2006, from http://www.thinker.org/gallery/index.asp.

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

Guy, M. and E. Tonkin (2006). "Folksonomies: Tidying up Tags?" *D-Lib Magazine* **12**(1). http://www.dlib.org/dlib/january06/guy/01guy.html

Hammond, T., T. Hannay, et al. (2005). "Social Bookmarking Tools (I): A General Review." *D-Lib Magazine* **11**(4). http://www.dlib.org/dlib/april05/hammond/04hammond.html

Hein, G. E. (1998). *Learning in the Museum*. New York, Routledge.

Janney, K. and J. Sledge. (1995, September). "'Formulating Typical Questions' in "A User Model for CIMI Z39.50 Application Profile"." Retrieved May 3, 2006, from http://www.cimi.org/old_site/documents/Z3950_app_profile_0995.html.

Jenkins, M. (2006a). Email Message: "everything i know isn't in the picture". to J. Trant, on Tue, 10 Jan 2006 08:56:49 -0500.

Jenkins, M. (2006b). Email Message: "Re: mma steve proof of concept data". to J. Trant, on Fri, 21 Apr 2006 16:50:44 -0400.

Jenkins, M. (2006c). Email Message: "timing in automated testing". to J. Trant, on Mon, 24 Apr 2006 15:35:39 -0400.

Jörgensen, C. (1999). "Image Indexing: An analysis of selected classification systems in relation to image attributes named by naïve users." *Annual Review of OCLC Research*. http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000002655:000000059275&reqid=17241

Jörgensen, C. and P. Jörgensen (2005). "Image querying by image professionals." *Journal of the American Society for Information Science and Technology* **56**(12): 1346 - 1359.

Leininger, K. (2000). "Interindexer consistency in PsycINFO." *Journal of Librarianship and Information Science* **32**(1): 4-8.

Markey, K. (1984). "Interindexer Consistency Tests: A Literature Review and Report of a Test of Consistency in Indexing Visual Materials." *Library & Information Science Research* **6**: 155-177.

Markey, K. (1988). "Access to iconographical research collections." *Library Trends* **37** (**2**): 154-174.

Mathes, A. (2004, December). "Folksonomies - Cooperative Classification and Communication through Shared Metadata." Retrieved August 6, 2005, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html.

McCorry, H. and I. O. Morrison (1995). report on the catechism project. Editor, National Museums of Scotland. Access 1995.

McKenna, G. and E. Patsatzi (2005). *SPECTRUM: The UK Museum Documentation Standard*. Cambridge, MDA.

Perrot, X., Ed. (2005). *ichim05: the international cultural heritage informatics meeting. proceedings*. Toronto, Archives & Museum Informatics.

Powerhouse Museum and S. Chan. (2005, ongoing). "Electronic Swatchbook." Retrieved January 10, 2006, from http://www.powerhousemuseum.com/electronicswatchbook/.

Quintarelli, E. (2005). *Folksonomies: power to the people*. ISKO Italy - UniMIB Meeting Milan. http://www.iskoi.org/doc/folksonomies.htm

Roberts, H. E. (2001). "A Picture is Worth a Thousand Words: Art Indexing in Electronic Databases." *Journal of the American Society for Information Science and Technology* **52**(11): 911-916.

Samis, P. (1999). "Artwork as Interface." *Archives and Museum Informatics* **13**(2): 191-198.

Shatford, S. (1984). "Describing a picture: a thousand words are seldom cost effective." *Cataloging and Classification Quarterly* **4**(4): 13-30.

Shirky, C. (2005). "Ontology is Overrated: Categories, Links, and Tags." *Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community [blog]* Retrieved April 26, 2006, from http://www.shirky.com/writings/ontology_overrated.html.

Shirky, C., S. Butterfield, et al. (2005). *Folksonomy, or How I Learned to Stop Worrying and Love the Mess*. O'Reilly Emerging Technology Conference San Diego, California. http://craphound.com/etech2005-folksonomy.txt

Shoham, S. and R. Kedar (2001). "The Subject Cataloging of Monographs with the Use of Keywords " *Cataloging & Classification Quarterly* **33**(2): 29-54.

Smith, G. (2004). "Folksonomy: social classification." *Atomiq / information Architecture [blog]*
Retrieved December 15, 2005, from
http://atomiq.org/archives/2004/08/folksonomy_social_classification.html.

Sormunen, E., M. Markkula, et al. (1999). *The Perceived Similarity of Photos: A Test-Collection Based Evaluation Framework for the Content-Based Image Retrieval Algorithms*. MIRA 99: Evaluating interactive information retrieval. Electronic Workshops in Computing, S. D. e. al., Ed.

Sundt, C. L. (2002). The Image User and the Search for Images. *Introduction to Art Image Access: Issues, Tools, Standards, Strategies*. M. Baca Ed. Los Angeles, CA, J. Paul Getty Trust.
http://www.getty.edu/research/conducting_research/standards/intro_aia/sundt.html

The Metropolitan Museum of Art (1968). *Computers and their potential applications in museums : a conference / sponsored by the Metropolitan Museum of Art, April 15, 16, 17, 1968*. New York, Published for the Museum by Arno Press.

The Metropolitan Museum of Art. (2005a). "Image Cataloguing Test Results. Fall 2004."
Retrieved May 3, 2006, from
http://www.steve.museum/reference/MMAImageCatalogingTestFall2004.pdf.

The Metropolitan Museum of Art. (2005b). "Image Cataloguing Test, December 7, 2004."
Retrieved May 2, 2006, from
http://www.steve.museum/reference/MMAImageCatalogingTest12-7-04.pdf.

The Metropolitan Museum of Art. (2005 -). "My Met Gallery."   Retrieved January 9, 2005, from
http://www.metmuseum.org/Works_of_Art/gallery.asp.

The Metropolitan Museum of Art. (2006). "Alvin Langdon Coburn, The Octopus, 1912, from The Metropolitan Museum of Art's Collections Highlights, Photographs."   Retrieved April 24, 2006, from
http://www.metmuseum.org/Works_of_Art/viewOne.asp?dep=19&item=1987.1100.13&viewmode=0&isHighlight=1.

The Metropolitan Museum of Art and K. J. Smith. (2005). "MMA Subject Cataloguing Trials."
Retrieved May 2, 2006, from
http://www.steve.museum/reference/MMASubjectCatTests0508.pdf.

Trant, J. and D. Bearman, Eds. (2006). *Museums and the Web 2006: selected papers from an international conference*. Toronto, Archives & Museum Informatics.
http://www.archimuse.com/mw2006/speakers/index.html

Trant, J. and B. Wyman (2006). *Investigating social tagging and folksonomy in art museums with steve.museum*. World Wide Web 2006: Tagging Workshop Edinburgh, Scotland, ACM.
http://www.steve.museum/index.php?option=com_weblinks&task=view&catid=35&id=32

Vander Wal, T. (2005). "Delicious Lesson and Social Network Ecosystems." *Off the Top [blog]*
Retrieved December 22, 2005, from
http://www.vanderwal.net/random/entrysel.php?blog=1765.

Victoria & Albert Museum, Ultralab, et al. (2005 -). "Every Object Has A Story."   Retrieved January 9, 2006, from http://www.everyobject.net.

Visual Resources Association. (2005, DRAFT - February 2005). "Cataloguing Cultural Objects: A guide to describing cultural works and their images."   Retrieved January 4, 2006, from
http://www.vraweb.org/ccoweb//index.html.

Visual Resources Association (VRA) Data Standards Committee. (2002, February 20, 2002). "VRA Core Categories, Version 3.0."   Retrieved January 4, 2006, from
http://www.vraweb.org/vracore3.htm.

Walsh, P. (1997). "The Web and the Unassailable Voice." *Archives and Museum Informatics* **11**(2): 77-85. also available at http://www.archimuse.com/mw97/speak/walsh.htm

10. About the Author

Jennifer Trant

A Partner in Archives & Museum Informatics, Trant is co-chair of the Museums and the Web and International Cultural Heritage Informatics Meeting (ICHIM) conferences, and has served on the program committees of the Joint Digital Libraries (JDL) and the Digital Libraries (DL) conferences, the Culture Program Committee of the International World Wide Web Conference, and the Board of the Media and Technology Committee of the American Association of Museums. She was Editor-in-Chief of *Archives and Museum Informatics: the cultural heritage informatics quarterly* (Kluwer Academic Publishers) from 1997-2000. Trant was the Executive Director of the Art Museum Image Consortium (AMICO) from 1997-2005. She has worked with the Arts and Humanities Data Service, King's College, London, England, and the J. Paul Getty Trust's Art History Information Program (for whom she managed the Imaging Initiative and directed the activities of the Museum Educational Site Licensing Project (MESL)). Trant managed disciplinary review and prepared the first edition of the report of the Art Information Task Force (AITF), entitled *Categories for the Description of Works of Art* for the College Art Association and AHIP. She has worked with computerized documentation systems at the National Gallery of Canada and the Canadian Centre for Architecture, Montréal. In addition to her consulting practice, Trant has taught Digital Cultural Heritage at the University of Toronto (Museum Studies and Information Studies) and is enrolled in the PhD program of the Faculty of Information Studies at the University of Toronto. A full vita and publications list can be found at http://www.archimuse.com/consulting/trant.html