

May 2015

# Preliminary Findings: A Comparative Study of Subject Metadata in an Images for Teaching Collection

Hannah M. Marshall  
*Cornell University*, [hmm88@cornell.edu](mailto:hmm88@cornell.edu)

Follow this and additional works at: <https://online.vraweb.org/vrab>

 Part of the [Classics Commons](#), [History of Art, Architecture, and Archaeology Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Marshall, Hannah M. (2015) "Preliminary Findings: A Comparative Study of Subject Metadata in an Images for Teaching Collection," *VRA Bulletin*:Vol. 41: Iss. 2, Article 5.  
Available at: <https://online.vraweb.org/vrab/vol41/iss2/5>

This Feature Articles is brought to you for free and open access by VRA Online. It has been accepted for inclusion in VRA Bulletin by an authorized editor of VRA Online.

---

# Preliminary Findings: A Comparative Study of Subject Metadata in an Images for Teaching Collection

## **Abstract**

In this study, art history and classics students were asked to perform descriptive tasks for art images from an Images for Teaching Collection. The descriptive terms that the participants assigned to the images were recorded and compared to the existing descriptive metadata for these images. Correspondence between the existing metadata and the participant-assigned subject terms was analyzed and characterized. Through this comparison, the researcher was able to determine approximate retrieval rates for subject based searching, analyze the types of subject analysis being done by each group, and test, through the use of a variable group, a potential framework for providing visual literacy outreach.

## **Keywords**

image indexing, subject analysis, subject access, image retrieval, descriptive metadata, subject metadata, controlled vocabularies, subject authorities, images for teaching, image collections

## **Author Bio & Acknowledgements**

Hannah Marshall is the Metadata Librarian for Image Collections at Cornell University. She has a background in art history and library science and her research interests include the subject analysis of images and works of art. Hannah was a 2014 IRDL Scholar and this study was accepted as part of the inaugural Institute for Research Design in Librarianship in June 2014.

## Introduction

The subject analysis of images is notoriously complex and the practice of cataloging and providing subject access to images is even more so. Image cataloging requires translating the non-verbal meaning of an image into language and then translating that language into a controlled vocabulary. These translations result in extremely low levels of inter-indexer consistency because they are not always intuitive and can be intellectually lossy. The subject analysis of images is subjective, and the resulting low levels of inter-indexer consistency result in subject metadata that is difficult to leverage in an automated environment. For these and other reasons, providing subject access to images tends to be done in a limited way, if at all.

At Cornell, the Images for Teaching collection enjoys the luxury of receiving full descriptive cataloging including subject access but, due to a lack of information about our users and their search behaviors, the search utility of the subject metadata for this image collection was utterly unknown. In an effort to address these gaps in our assessment and improve our grasp of users' needs and expectations, the Metadata Services unit of the library undertook a study wherein users were asked to describe a set of images by assigning subject terms to each image. Their responses were compared to the library's existing metadata for each image. Through this comparison, the researcher was able to determine approximate retrieval rates for subject-based searching, analyze the types of subject analysis being done by each group, and test a potential framework for providing visual literacy outreach.

In October of 2014, a survey was circulated among all undergraduate students at Cornell who were enrolled in either an art history or a classics course, in which participants were asked to assign subject terms to a set of ten images pulled from the Images for Teaching collection. The collection is used primarily for teaching by faculty in the College of Arts and Sciences – primarily the departments of Art History and Classics. Development of the collection is driven most heavily by faculty from these departments, so the target population for the study was determined to be students taking courses in these departments based on the assumption that they would correspondingly be the heaviest student users of the images.

The Images for Teaching collection includes 44,000 images of over 30,000 individual works. Images are cataloged in a local database using VRA Core 4.0<sup>1</sup>, Cataloging Cultural Objects (CCO)<sup>2</sup>, and controlled vocabularies from the Getty<sup>3</sup> and Library of Congress<sup>4</sup>. As current practice, ref IDs (reference IDs) for authorities are captured whenever available for use in current and future discovery environments.

Subject cataloging is performed for every image in the collection; image catalogers within the library's technical services department consult an image's source material and perform additional research when necessary in order to capture descriptive and subject metadata for each image. Due to the variety of material, the challenging nature of the source material, and personnel limitations, this work can be extremely time- and resource-intensive. In conjunction with a desire to learn more about the way these collections are serving our users, this project offered a good opportunity to evaluate our practice of providing full subject access to all images in this collection and think critically about how best to serve the users, the collection, and the catalogers.

---

<sup>1</sup> <http://core.vraweb.org/>

<sup>2</sup> <http://cco.vrafoundation.org/>

<sup>3</sup> <http://www.getty.edu/research/tools/vocabularies/>

<sup>4</sup> <http://www.loc.gov/library/libarch-thesauri.html>

## Research Questions and Methodology

The study's design was driven by three research questions, each of which drove a specific aspect of the data collection and analysis.

*What is the level of correspondence between the existing subject terms for these images and the participant-assigned subject terms?*

The first question was focused on determining the levels of correspondence between the literal terms in the survey responses and subject terms in the existing metadata. In doing this comparison, an instance of correspondence was regarded as a proxy for successful image retrieval in a subject search. Similarly, an instance of non-correspondence was considered an example of an unsuccessful image retrieval. Using this approach, we compared the existing terms for each image to the terms in the survey responses and determined the percentage rates at which an existing term would have yielded successful and unsuccessful image retrievals.

*What is the level of correspondence in the types of subject terms assigned by users and those in the existing metadata?*

The second question was designed to expand upon the findings of the first by looking at the terms that yielded successful and unsuccessful image retrievals and trying to account for that success or failure. At the root of this question was the suspicion that instances of unsuccessful image retrieval were due, at least in part, to one or a combination of two things: First, the participants and the image catalogers were likely to use different terms for describing the same concepts as a result of natural incompatibilities between free text search and controlled vocabularies. Second, the two groups might differ in how they analyzed the subject of an image, creating a situation where each group not only uses different terminology for expressing identical concepts, but they found fundamentally different concepts to be significant in understanding and capturing the meaning of an image.

The second research question was designed to lend some insight into the types of terms being assigned to an image by each group as a way of evaluating the general similarities and differences in the subject analysis being done by each group. To do this, a system for categorizing subject terms into definitive types had to be designed and each term then had to be coded according to its type. Relative numbers of each type of term assigned by each group could then be compared.

To develop a system for categorizing terms, literature was reviewed dealing with both the subject analysis of images and works of art, because the collection consists almost entirely of images of works of art. The system needed to be robust enough to handle every term in the result set and systematic enough that different people would make the same decisions about the category to which a term should be assigned.

Ultimately, a tripartite system heavily indebted to Erwin Panofsky's *Studies in Iconology* was developed in which each tier is defined by the type of question a term answers about the image to which it is assigned. So, terms assigned to the 'Primary' level are those which answer the questions, "What is this image of?" or "What does this image include?", addressing a topical level of analysis wherein objects and elements are identified and named. Terms assigned to the

'Secondary' level are those which answered the question, "What is this image about?" At the secondary level, the objects and elements are interpreted; characters are identified, facial expressions and emotional states are asserted, and gestures are ascribed meaning. Terms assigned to the 'Tertiary' level are those which answer the questions, "What is the image a good example of?" or "How does the image communicate?" Tertiary level subject analysis has traditionally been conceived of as a synthesis of the primary and secondary levels infused with an awareness of the work or image as an expressive output and a product of a time and place. It is to this level that terms identifying artistic devices and sophisticated systems of signification belong.

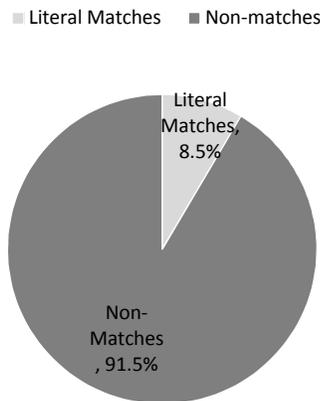
*Does providing users with a framework for analyzing the subject of an image change the nature and content of the subject terms they choose to assign to that image?*

The third research question drove the design of a variable group in which a framework for providing visual literacy outreach was tested. The questions that were used to define the categories in the second question were reintroduced in the version of the survey circulated to the variable group. Above each image in the variable version of the survey were the three questions: What is this image of? What is this image about? What is this image a good example of?

The variable group was designed to yield insights into the users and their search behavior. First, if it was found that the variable group's responses exhibited a significantly higher level of correspondence with the existing metadata, that would be compelling evidence for using those questions as a framework for visual literacy outreach to students. Second, if no significant difference was detected, it is likely that a systematic approach to subject analysis paralleling these three questions is already in use by the participants, indicating that this type of analysis is either intuitive or a skill that is quickly absorbed by students enrolled in art history and classics courses.

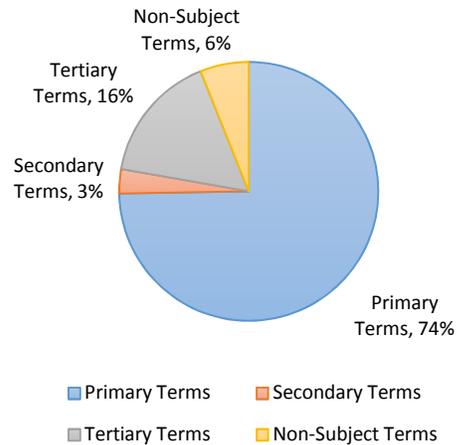
## Findings

**Literal correspondence between responses and existing metadata**



**Figure 1.1**

**Corresponding literal terms broken down by type**

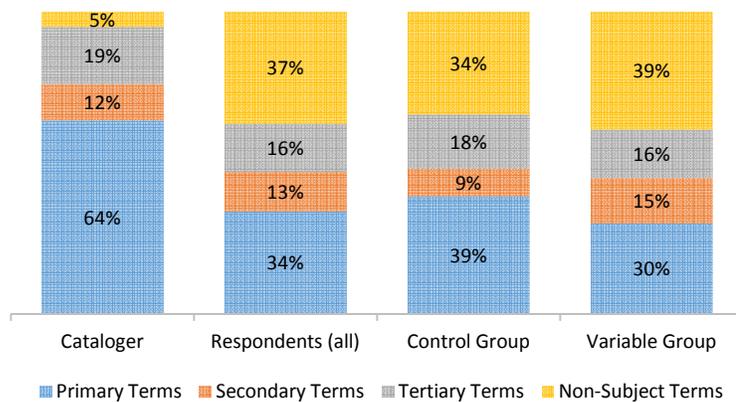


**Figure 1.2**

The first comparison was between the participants’ responses and the existing metadata for each image in which the literal terms were compared in order to get a sense of the retrieval rates for subject searches. As shown in Figure 1.1, the rate of successful image retrieval was 8.5%, meaning that for only 8.5% of the terms in the existing metadata was there an instance of correspondence among the participants’ responses.

When that 8.5% correspondence was analyzed according to term type, it was discovered that, of the terms which yielded successful image retrieval, 74% were primary terms or those answering the questions, “What is this image of?” or “What does this image include?”

**Comparison of all types of terms assigned by each participant group**



**Figure 2.1**

The survey responses and the existing metadata (identified in Figure 2.1 as “Cataloger”) were coded according to their different types for comparison. Included along with “primary”, “secondary”, and “tertiary” is the category, “non-subject terms” which are exactly

that: descriptive terms which do not address the subject meaning of the image. These terms often address aspects of an image that are captured in other fields within the descriptive metadata.

Immediately evident in Figure 2.1 is the high percentage of non-subject terms assigned by participant groups relative to the existing metadata. Similarly, the existing metadata includes nearly double the percentage of primary terms that the participant groups use to describe the same images.

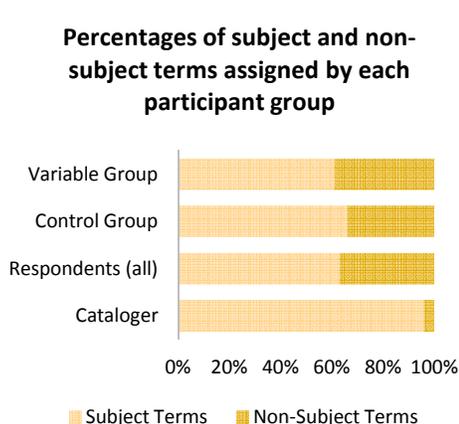


Figure 2.2

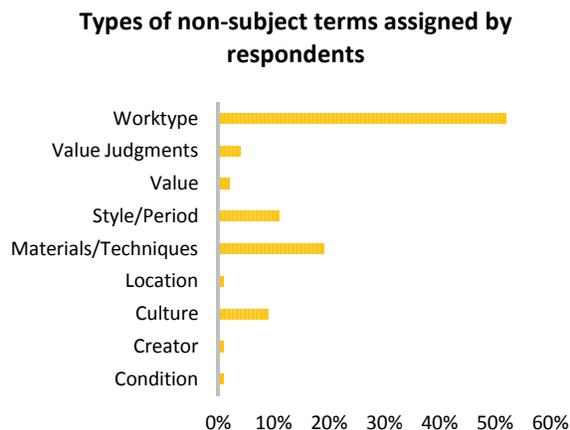


Figure 2.3

Approximately 75% of the non-subject terms assigned by the participants were terms describing physical properties of the work depicted in the image, like the worktype or the materials and techniques used in the work’s production. The next most significant type of non-subject terms were those addressing information about the style or period of the work and the culture responsible for producing it.

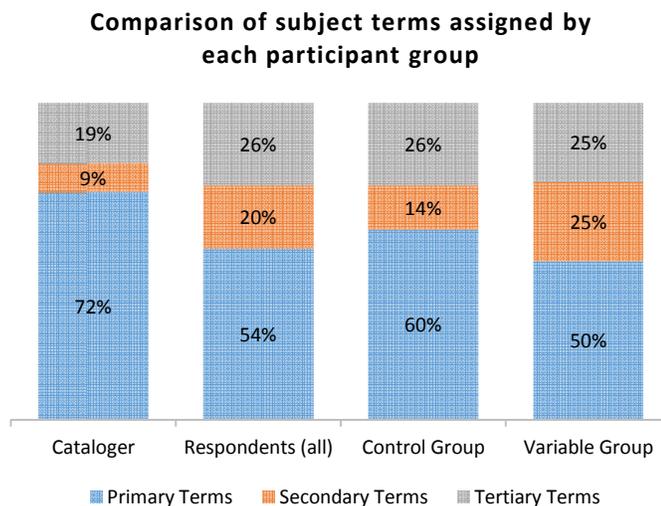


Figure 2.4

When Figure 2.1 is adjusted to exclude the non-subject terms, it becomes very clear that the rates at which each type of term appears in the survey responses and the existing metadata are similar. The existing metadata is 72% primary terms whereas the participants’

responses include a more diverse mix of term types, but both groups are still over 50% primary terms and less than 30% tertiary terms.

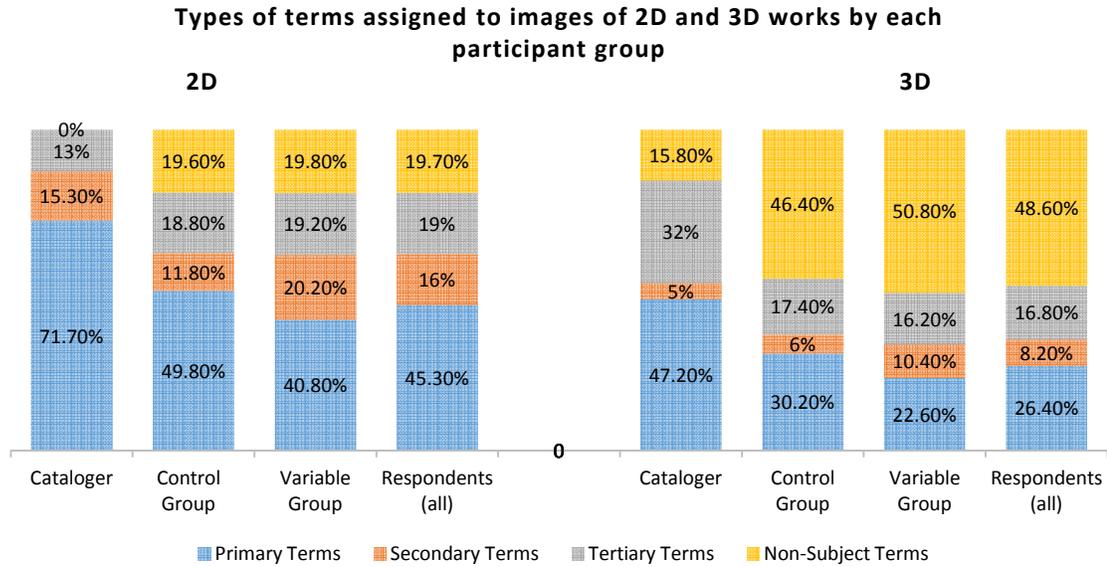


Figure 2.5

After drawing comparisons between the existing metadata and the participants’ responses, the responses were broken down by image to look for patterns and outliers at the image level. Figure 2.5 captures a pattern in the types of terms assigned by the participants to images of 2D works like paintings, and images of 3D works like vases or jewelry. Visible in this figure is the significantly higher rate at which non-subject terms were assigned to images of 3D works relative to images of 2D works. Among participants, these rates were more than twice what they were for images of 2D works, and for the existing metadata (identified once again as “Cataloger” here) the only instances of non-subject terms were in the description of images of 3D works.

**Participant-assigned primary and non-subject terms assigned to 2D and 3D works**

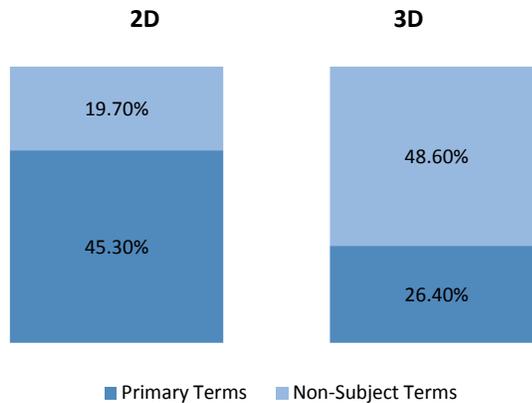


Figure 2.6

Upon closer inspection in Figure 2.5, it becomes clear that there is a relationship between the application of primary and non-subject terms and how that application differs between images of 2D works and 3D works. As seen in Figure 2.6, primary terms are assigned to images of 2D works at roughly the same rate that non-subject terms are assigned to images of 3D works. The ratio of primary terms to non-subject terms for each type of image is nearly the exact inverse of the other, suggesting that the participants applied non-subject terms to images of 3D works in the same way they were applying primary terms to images of 2D works, and vice versa.

The significance of this will continue to be explored in future rounds of data collection. It seems reasonable that this relationship reflects a failure among the participants to distinguish between the content of a work and the content of an image, which caused them to treat the content of a 2D painting captured in an image the same way they treated a 3D vase captured in an image. More generally, this may indicate a visual literacy threshold where analytical training, like the distinction between a work of art and an image of a work of art, does not get translated into information-seeking behaviors.

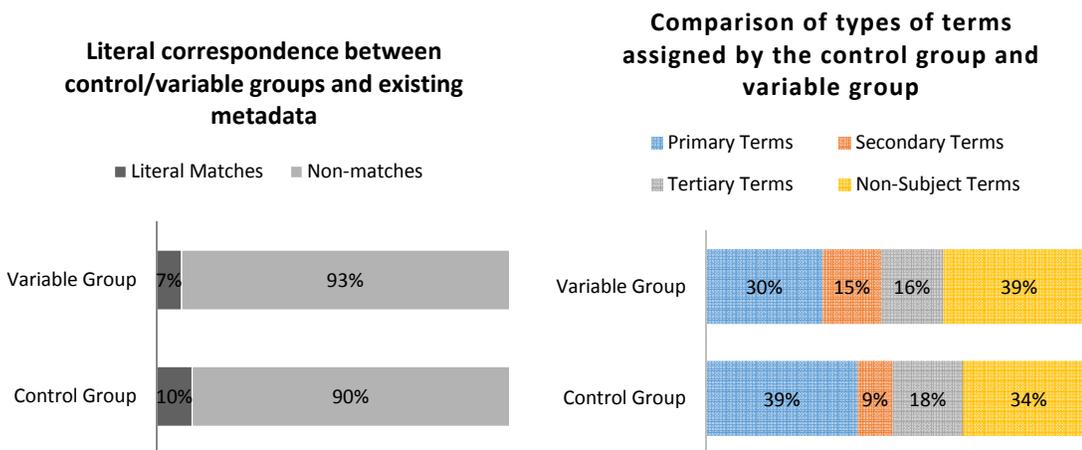


Figure 3.1

Figure 3.2

Comparisons of the control and variable groups revealed very little difference in the nature and content of the responses between the two groups. The variable group had slightly lower retrieval rates than the control group – 7% compared to 10% - and the two groups differed slightly in the ratio of term types they assigned. The variable group assigned 9% fewer primary terms than the control group and 5% more non-subject terms; given the search utility of primary terms evidenced in Figure 1.2 and the irrelevance of non-subject terms in a subject search scenario, these two differences may be largely responsible for the lower retrieval rate among the variable group.

## Discussion

The results yielded four key findings, two of which have very clear applications in image cataloging practice if they are borne out in subsequent rounds of data collection.

1. *Primary terms yield the greatest search utility and higher levels of successful image retrieval*

Based on the extremely high rate of primary terms among successful image retrievals, it is clear that primary terms contribute greatly to the search utility of an image's subject metadata. Considered alongside the finding of Figure 2.5 that primary terms were used most heavily to describe images of 2D works, a logical application of this conclusion might involve shifting image cataloging practice to focus on primary level cataloging for images of 2D works.

2. *High numbers of non-subject terms applied to images of 3D works suggest that subject metadata is a weak access point for images of 3D works*

Whatever the reason for the high rate of non-subject terms assigned to images of 3D works, the preference for non-subject access points for these images is clearly borne out in the survey results. Whether this can be attributed to a conflation of the work/image relationship by the participants remains to be seen, but it is clear that approximately 75% of the non-subject terms in the responses would likely have yielded successful image retrievals in a basic keyword search. It is also worth noting that the overall search utility of these images is improved when the non-subject terms are considered in the context of a general keyword search rather than a subject search, and the former is the default search mode in this collection, and presumably most collections of images.

To leverage this finding into more effective image cataloging, it may make sense to forego subject cataloging for images of 3D works, focusing instead on providing other descriptive access points – most significantly those addressing the physical properties of the work such as its worktype, the materials and techniques used in its creation, and its condition.

3. *Poor image retrieval rates seem to be due to incompatible vocabularies between the participants and the image catalogers, rather than radically different approaches to subject analysis*

This is most clearly captured in Figure 2.4 which reveals that the ratios of each type of subject term assigned by the participants and the image catalogers are very similar. When this is considered in the context of the high rates of unsuccessful image retrieval captured in Figure 1.1, it is clear that the non-correspondence responsible for those rates must be happening at the term level rather than the analysis level. This is made even more likely given the use of controlled vocabularies in the production of the existing metadata; controlled vocabularies are not designed to align with natural language, much less in a search scenario that has been shaped in recent years by commercial search engines and the simplicity of single-search box interfaces.

4. *Including a framework for analyzing the subject of an image did not change the nature and/or the content of the participants' responses*

The lack of significant difference between the control and variable groups, in conjunction with the finding in Figure 2.4 that the participants and the image catalogers assigned various types of subject terms at similar rates, suggests that the framework offered to the variable group reflects an intuitive approach to subject analysis that is already in use among the participants. In short, the variable group was relatively unaffected by the coaching because they were being told how to do something they already knew how to do.

### **Conclusion**

A second round of data collection for this project is currently under way and these preliminary findings will be adjusted to include that data set. There are a number of areas for further analysis using the existing data, and adjustments to the study design that could potentially yield interesting insights.

Areas for further analysis using existing and incoming data are diverse. While the design of the study was simple and data collection involved only a modest time commitment, the data yielded by the study has proven to be rich and varied, responding to a number of different types of analyses.

In order to more thoroughly explore the theory that low retrieval rates are likely due to incompatible vocabularies, the researcher will look at synonyms between the two data sets and explore the rates of concept-correspondence between the groups. Identifying specific incidences of unsuccessful image retrieval exhibiting conceptual correspondence will increase the certainty with which we can diagnosis and correct low retrieval rates in image collections.

Another avenue for further analysis is assessing literal matches according to the degree of the match rather than as a binary value. There are certainly instances where a term in the existing metadata appeared multiple times in the participants' responses, and a closer investigation of the terms (and types of terms) that yielded strong matches would be a meaningful addition to our understanding of overall search utility.

In the second round of data collection, additional sources of metadata will be explored in order to examine inter-indexer consistency and reduce potential anomalies within that data set. The design of the study may also be adjusted in future rounds of data collection to mimic a faceting-based search environment rather than a free text one to determine differences in the retrieval rates between the two scenarios.

This study has proven to be a powerful exploratory assessment tool lending insight into users and their search behavior, current image cataloging practice, and the search utility of the collection. Potential avenues for applying the findings are clear and simple and, through continued data collection and analysis, the Metadata Services unit will continue to improve discovery and access for this and other image collections.