

Constructing Descriptive Records for an Art Image Database: What Do Use Statistics Tell Us?

Peter Hepburn and Joan B. Fiscella

The study compares three sample sets of records taken from the AMICO database to examine possible factors influencing retrieval of images: named artist and artist reputation, word count, and record richness. The authors found that images of works by renowned artists tended to show high numbers of retrievals. When works depicted were by relatively unknown or anonymous artists, more retrievals were likely if accompanying records included higher unique word counts. The frequency of first occurrences of name, geographic, and time terms in the records showed no major differences among the three sets. The authors suggest a strategy for constructing image records.



Libraries today contribute to the content of digital space. A library typically provides access to its catalog, publishes information about itself and its collections, and extends its services to users outside its physical site. Additionally, a library may digitize portions of its own collections or public records of its parent institution for user access. Images of artworks or other objects in all media are a fertile area to be made available. Digi-

tization allows virtual use by those who cannot travel to view the original works. Libraries may display online fragile objects for those who do not need immediate contact with the original. Without good means of access, though, scholars and other users cannot find what has been digitized.¹ High-quality descriptions attached to the digitized images may increase accessibility and thus potential use; therefore, it is worth examining retrievals from an existing image database to learn

Peter Hepburn is Assistant Circulation Librarian and an Assistant Professor in the Richard J. Daley Library at the University of Illinois at Chicago; e-mail: phepburn@uic.edu. Joan B. Fiscella is Bibliographer for Professional Studies and an Associate Professor in the Richard J. Daley Library at the University of Illinois at Chicago; e-mail: jbf@uic.edu. The authors are grateful to Hava Kagle for retrieving the records in their random samples from the AMICO database and to the Research Libraries Group (RLG) for its permission for Ms. Kagle's work. They acknowledge the assistance of Milton Beope and Aroind Kuppan in downloading records, conducting word counts, and searching the OCLC database, with thanks to the UIC Library's Faculty Development Advisory Committee who provided funding for student assistants. They also acknowledge former colleague, Emily Bounds, who helped with the early literature searching, Conrad Paulson for his help with the statistical analysis, and the manuscript's reviewers for their suggestions. The authors thank the Museum of Fine Arts, Boston, and the Library of Congress Prints and Photographs Division, Washington, D.C., for permission to use its records. Finally, they thank colleagues Deborah D. Blecic, Stephen E. Wiberley Jr., and John M. Cullars (who pointed out the potential problem with determining photographers' reputations) for their perceptive comments and suggestions.

what characteristics of description are associated with higher use.

Use statistics for databases assist in assessing the relative value among electronic resources, help to justify selection, and serve as an indicator of the value of these collections to the provider's clientele. Although use statistics are not the sole or even a sufficient measure of resources' values due to current limitations of use statistics,² they are indicators. High usage generally indicates that the database or group of databases is providing information needed by a greater number of users than are low-use databases. Within a particular indexing database, use statistics also might indicate, for example, which of the full-text journal articles are retrieved more often than others, thus informing collections development decisions. Use statistics may point to the need to reposition certain resources on Web pages or the need to incorporate one or another database into instructional activities or materials. In a database where images of artifacts are available, use statistics might inform selection decisions for future digitization efforts. These statistics also might contribute to understanding what kind of information in an accompanying record is important to a potential user.

To gain understanding of access factors in the construction of an image database, the authors studied one year's use statistics of the AMICO database. Specifically, they examined academic institutions' usage for fiscal year 2000–2001 (the most recent available at the time of the study) and the descriptive records accompanying images of artworks retrieved during that period.

About AMICO

The AMICO (Art Museum Image Consortium) database was a collaborative project that ran from 1997 to 2005, with academic institutions and museums contributing records. It consisted of images depicting works of art in collections mostly in the United States and Canada. At the time of this study, the Research Libraries Group

(RLG) interface permitted simple and advanced searching techniques that retrieved thumbnail views of the resulting images. Clicking on a thumbnail pulled up a larger view of the image accompanied by the written description of the item, including the physical characteristics of the original work, the history of the item, its media type, and provenance. Two other views of each image were available: magnifications of the original images unaccompanied by any description.

Underlying the displayed description of each image was a structured record. (See figure 1.) AMICO's 1999 Data Specification Manual provided the requirements and guidelines for constructing the record. The catalog fields combined and translated into a limited number of headings for the display: creator, work, ownership, commentary, descriptive terms, history, and context. Each record required creator, work, and ownership information; entries under other headings were optional.

The AMICO database offered two levels of searching: simple and advanced. A user could perform any one of three simple types: creator, title, or keyword. Creator searches referred to searches for the maker of the original work, whether a specific person, a school of art, or a cultural background. Examples included "Pablo Picasso," "follower of Michelangelo Buonarroti," and "Western Mediterranean." Title searches referred solely to the title of the work of art depicted. Keyword searches crossed all indexes. Multiterm searches defaulted to combining terms with an implied "AND." No explicit Boolean operators were permitted in the simple keyword search.

Advanced searching offered a greater number of search fields: creator, title, type (category of objects, such as drawing or furniture), materials/technique (medium or media used), date (an approximate range or specific date), owner name (the institution or person owning or holding the work), owner place (location of owner), and ID (an accession number

assigned to each image according to an alphanumeric scheme devised by AMICO that built on the institutions' numbering). The advantage of these search fields was greater specificity of searching. The interface also allowed combining searches, thus enhancing specificity. Consequently,

the user could search for terms in one index and combine the search with another in the same or a different index using a choice of three Boolean operators. Both simple and advanced searches permitted index browsing for all fields other than keyword.

FIGURE 1
Two Examples of AMICO Image Record Descriptions,
Showing Extremes of Categories of Data Included

<p align="center">^aCreator</p> <p align="center">Artist not recorded</p> <p align="center">Work</p> <p align="center">3 Etchings, Dates not recorded Unmeasured Books MATERIAL NOT RECORDED</p> <p align="center">Ownership</p> <p align="center">Museum of Fine Arts, Boston, Boston, Massachusetts, USA Bequest of W. G. Russell Allen Rights Accession Number: 63.752</p>	<p align="center">^bCreator</p> <p align="center">Arnold Genthe 1869–1942 Photographer</p> <p align="center">Work</p> <p align="center">Jordan, Miss, in Bush's garden, Title devised from Genthe's records title, 1913 Sept. 5 5 x 7 in Photographs, 1 Nitrate negatives</p> <p align="center">Ownership</p> <p align="center">Library of Congress Prints and Photographs Division, Washington, D.C., USA Courtesy of The Library of Congress Rights Accession Number: LC-G401-T01-0415-E</p> <p align="center">History</p> <p align="center">Provenance: Genthe Estate; Purchase; 1942 or 1943</p> <p align="center">Commentary</p> <p align="center">Context: Additional annotation: in Bush's gardens</p> <p align="center">Related Materials</p> <p align="center">Works: Arnold Genthe Collection (Library of Congress). Negatives and transparencies</p> <p align="center">Descriptive Terms</p> <p align="center">Negative Nitrate Gardens</p>
--	---

Sources: Artist not recorded, [3 Etchings], Museum of Fine Arts, Boston, Boston, MA, 63.752, The AMICO Library BMFA.63.752; Arnold Genthe, Jordan, Miss, in Bush's garden, Library of Congress Prints and Photographs Division, Washington, D.C., LC-G401-T01-0415-E, The AMICO Library LOC_age96003420/PP.

^a Record drawn from HighR Set as an example of a description using the minimum of three categories (headings) of data

^b Record drawn from MidR Set as an example of a description using the maximum of seven categories (headings) of data

Method of Investigation

This study analyzes the use statistics of the AMICO database for the fiscal year 2000–2001 to determine what factors might contribute to higher use of some records and lower use of others. The authors examined only records that had been retrieved by users of the database and only those retrievals attributed to academic institutions. Because the AMICO system did not retain all instances of data showing one or two retrievals during 2000–2001, the study’s population of 46,419 records included only those showing three or more retrievals.

For purposes of comparison, the data set of records retrieved three or more times was divided into three distinct groups that the authors characterized as high retrievals, low retrievals, and mid-retrievals. The authors distinguished the sets in two ways. The high-retrieval set stood apart by a break in the number of retrievals; it consisted of 74 records with 800+ retrievals. The low-retrieval set of 1,215 records also stood apart, but by virtue of having the minimum number of retrievals for which there were complete data. The mid-retrieval set consisted of the remaining 45,130 records with 4,514 retrievals. Based on the sampling guidelines of Robert V. Krejcie and Daryle W. Morgan,³ the authors drew 63 records from the high-retrieval set. From the original mid-retrieval set of 45,130 records they drew 381; from the original low-retrieval set of 1,215 records they drew 292.⁴ The authors refer to these sample sets as HighR, MidR,

and LowR Sets, respectively. Of the 63 records in the HighR sample, 62 were usable.⁵ Of the 381 records in the MidR sample, 363 were usable. Finally, of the 292 records in the LowR sample, 290 were usable. (See table 1.) With the cooperation of RLG, the authors contracted with an RLG employee who provided the full catalog records of the samples.

For this study, the authors considered three factors in relation to use levels: reputation of creators of the works, the word count in records accompanying images, and the richness of language used in the descriptions.

Artist Reputation

The artist’s name is an important factor in users’ access to records in humanities resources, as research has shown. (Although the AMICO record heading refers to “creator,” the authors will use “artist,” reserving the former term for more generic uses.) In his first study of precision of humanist’s vocabulary, Stephen E. Wiberley Jr. examined the terms used as entry points in humanities reference works. He called the “singular proper term,” that is, the name of a person or single creative work, the most precise proper term (“proper” referring to one of a class of things).⁶ Even two or more people or multiple works having identical names or titles, respectively, can meet that standard of precision thanks to geographic and time designations that help to distinguish them from one another. He showed in his sample of entry points that

TABLE 1
Composition of AMICO Record Sample Sets

Sample set	Range of Retrievals per Record	Records			Named Artist Records (% of Sample)	Named Artists
		Database	Sample set	Usable sample		
HighR	800–1,232	74	63	62	60 (97)	4
MidR	4–514	45,130	381	363	305 (84)	187
LowR	3	1,215	292	290	253 (87)	92

58 percent were singular proper terms. Although Wiberley refined his analysis in a later study, he reaffirmed the importance of the names of creators of literature, music, and the arts.⁷ More recently, Linda H. Armitage and Peter G. B. Enser studied users' requests for images in seven picture libraries.⁸ Their findings put little emphasis on requests by named artist in each of six of the image libraries (the focus ranging from local photographic history, locomotive archives, to aerial photographs); the requests by named artist were under 3 percent. In contrast, the seventh of this group, a library affiliated with an academic institution with collections by renowned photographers, showed almost 11 percent of its requests by named artist. This library, too, had the highest percentage (43.5%) of requests by known items. However, the study did not indicate whether known items included reference to the photographers' names.

The authors of this study investigated the role of personal names in the use of AMICO records. They asked whether, in the randomly sampled data, the records with personally named artists show a greater number of retrievals (or higher use) than those records with unknown, group, or culturally designated creators. Furthermore, they investigated whether the reputation of the personally named artist influences the amount of use of the records.

As noted above, the original data of the present study were divided into three sets, grouped by the number of retrievals, that is, the level of usage. Thus, the first indicator of the importance of the personal name is the difference among the usage samples in the number of personally named artists (that is, known artist or attribution to the artist, but not studio of, nor culture, nor unnamed creators)⁹ responsible for the work captured in the image and its accompanying record. Of the HighR Set of 62 records, 60 (97%) indicate personally named artists; of the MidR Set of 363 records, 305 (84%) have personally named artists; and of the LowR

Set of 290 records, 253 (87%) records show personally named artists (table 1). Thus, each of the three sets includes a large percentage of records with named artists, with the HighR Set exceptionally large.

Just as the number of records with named artists varies by set, so does the number of unique artists in each of the sets. The HighR Set of 62 records includes works by just four personally named artists whereas the MidR Set of 363 records includes works by 182 personally named artists and the LowR Set of 290 records, 92. In the HighR Set, three of the four artists are represented by 19 or 20 records each, accounting for more than 19,000 or 20,000 combined retrievals of all records for each artist whereas one artist is represented by only one record, accounting for more than 1,100 retrievals. The four names account for a mean of 15 records each and a mean combined retrieval of 1,016 per record. In contrast, the total number of records for personally named artists in the MidR and LowR Sets (removing the Picasso records, the only artist duplicated from the HighR Set) comes to 554 with combined total retrievals of under 12,000.

More significant is the reputation of the named artists. The authors investigated the hypothesis that the records with the greatest number of retrievals were those with images of works done by artists who are the most renowned. To determine the relative reputation of the creators of works with images and accompanying records included in the study sample, the authors compared the number of monographs published about each artist.¹⁰ They assumed that a prolific and/or influential artist would be the subject of a great many monographic works. They used the number of records in the OCLC catalog as an indicator of the amount of publishing done about the creator and used the number published as a surrogate for the artists' reputations relative to one another. They performed a subject search of each artist's name in the OCLC catalog (WorldCat, in FirstSearch); all the searches were done in one day (October 24, 2003)

to minimize the possibility of artificially weighting names for which records had been added at later dates. They counted the total number of book records associated with each artist's name. The results of these searches showed that personally named artists in the HighR Set are subjects of a great many books. The average number of records for book materials in WorldCat for these four artists was 1988.75, as compared to an average of 153.08 for personally named artists in the MidR Set and 34.91 for those in the LowR Set. (See table 2.) Thus, the pattern of artist reputation follows the pattern of usage in the sample sets.

Although there are artists with large numbers of book records about them appearing in the MidR and LowR Sets of records, the proportion is quite different from the numbers in the HighR Set. The number of book records in WorldCat

about the four artists in the HighR Set ranged from 876 to 3,665. In the MidR Set, 145 of the 182 artists (including Picasso, the one artist who appears in both HighR and MidR Sets) were subjects of book records in WorldCat with the number of found (non-zero) book records ranging from 1 to 3,665 (the number for Picasso was 3,665; the next highest in the set was Rembrandt Harmensz van Rijn with 2,284 book records). In the MidR Set, each of fifteen (8.2%) personally named creators of original works captured in AMICO images had more than 500 book records retrieved in WorldCat whereas only two (3.3%) of the 60 named artists with records in WorldCat in the LowR Set had more than 500 book records retrieved in WorldCat. The number of found book records about those artists in the LowR Set ranged from 1 to 755 (Pierre Auguste Renoir). (Artists whose records appear in

TABLE 2
Mean WorldCat Records per Creator, Mean AMICO Records per Creator,
and Mean Retrievals Per AMICO Record

	HighR		MidR		LowR	
Number of named artists	4		187		92	
	Mean	Range	Mean	Range	Mean	Range
<i>WorldCat records per named artist</i>	1,988.75	876–3,665	153.08	0–3,665	34.91	0–755
<i>AMICO records per named artist</i>	15.00	1–20	1.72	1–50	2.75	1–114
<i>AMICO retrievals</i>						
All records	1,015.87	800–1,232	39.96	4–514	3.00	3
Named artist records	1,015.82	800–1,232	39.01	4–514	3.00	3
Named artists with zero WorldCat records	N/A	N/A	34.05	4–358	3.00	3
Unnamed creator records	1,017.50	987–1,048	44.93	4–294	3.00	3
Unnamed creators and artists with zero WorldCat records	1,017.50	987–1,048	40.49	4–358	3.00	3

TABLE 3
Pearson Product Moment Correlations of WorldCat Records to AMICO Retrievals

Subset	Coefficient		
	HighR	MidR	MidR w/o Genthe Records
All named artist records	-0.183	0.479*	0.461*
Named artist records (>0 WorldCat records)	-0.183	0.517*	0.493*

* Statistically significant results
 Note: For all correlations in the LowR Set, no coefficient was returned as each record was retrieved the same number of times.

both the MidR and LowR Sets were not removed from either.)

It is clear that the highest number of retrievals of the AMICO study samples is for images of works done by artists who are well known or influential as indicated by the number of book records in the WorldCat database. The authors of this study sought to refine the results by investigating whether a relationship of retrievals to artist reputation also holds *within* each of the three sets of randomly selected records and whether the relationship is statistically significant. To determine significance, the authors ran the Pearson Product Moment statistical test for each of the three sets. They used this test to determine the correlation between two sets of values from each of the three sample sets. Tables 3 and 4 show the results of the statistical testing. In the HighR Set, only two records had no named artist and the four named artists are highly prominent. Working with so few data yielded a negative correlation between artist reputation and record retrieval, though not one that was significant at a level of 0.05. Because all records in the LowR Set had the same number of retrievals, no correlation coefficient was returned.

With its larger population and wider range of data, the MidR Set showed the most variability of the three sets in terms of number of personally named artists and of retrievals of records. For all named artists in the MidR Set, a comparison of

the number of WorldCat records for each artist with the total number of retrievals of image records by that artist (table 3) yielded a correlation coefficient of 0.479, significant at the 0.05 level. Furthermore, when the correlation is run using only named artists with at least one monographic record in WorldCat, the coefficient strengthens (0.517), demonstrating greater significance at the 0.05 level compared to all named artists. Thus, the authors concluded that there is an indication in the large and varied MidR Set that artist reputation influences level of usage of records; this conclusion corroborated the finding that artist reputation helps account for the levels of use among all three sets.

If artist reputation were the only variable that could account for the difference in usage of image records, one would expect that the artists' names would appear in one or another set, but not more than one set. However, there are a limited number of artists whose image records appear in two sets. For example, records of the works of Picasso appear in both the HighR and MidR Sets. There is greater overlap between the MidR and LowR Sets with fifteen creators appearing in both. This suggests to the authors that, unsurprisingly, there is a varying level of interest among works of even the well-known creators. The authors did not investigate whether there is a statistical relation between reputation of particular works and their retrieval.

Arnold Genthe

Among the records the authors examined were a surprising number with Arnold Genthe named as artist, raising questions about whether the sample was skewed. An early twentieth-century photographer, Genthe accounted for 114 (39.04%) of the 290 records in the LowR Set. In the MidR Set, Genthe accounted for 50 (13.78%) of 363 records. No other single artist is responsible for so many images in any of the sample sets. Of the 715 records in the three sample sets, 211 (29.51%) were contributed by the Library of Congress, including all 164 records of works by Arnold Genthe.

Genthe, with just 14 subject WorldCat records, was not among the most renowned artists in the sample. The use data showed that records of works about Genthe from the MidR Set were retrieved an average of 11.4 times, with a range of 4 to 61 retrievals. In fact, there is a large gap between the second-most-retrieved record, with 28 retrievals, and the most, with 61 retrievals. In general, Arnold Genthe may be characterized, at least by the criteria used in this study, as a lesser-known artist whose works were not retrieved especially often.

The prevalence of Genthe records may be easily explained. As Henry Pesciotta reports,¹¹ in the early days of the AMICO database, only a few institutions, including the Library of Congress, contributed their records. Original member institutions relied on records that had already been created for special projects or events rather than on records deliberately chosen to highlight either their best-known or most obscure collections. Thus, the large number of Genthe records in the sample may indeed be representative of their presence in the database.

The presence of the large number of Genthe records also may indicate the limitation of the use of WorldCat book records as a criterion of reputation for some artists. In this case, that criterion portrays Arnold Genthe as someone less known or less esteemed than other artists

in the sample. A different view is suggested by Shaw's *A Century of Photographs 1846–1946*.¹² In the foreword, Curator of Photography Jerald C. Maddox notes that the

work of a particular photographer like Arnold Genthe is represented in all its aspects, from negatives to studio proofs to finished exhibition prints. Such a collection offers a unique opportunity to study several stages of photography as a means of personal expression (p. viii).

Further, Paul Vanderbilt's essay accompanying the photographic prints selected from Genthe's work describes the labor-intensive work in repairing, selecting, and conserving the negatives; he says:

Genthe ... eminently deserves this degree of care in the national collection. As a technician, he did much to accomplish the revolution in photography which ... gave to the art its present outstanding position (p. 86).

Thus, while the number of WorldCat book records works as a criterion for the reputation for many artists, including some photographers, it is not adequate in the cases of all photographers.

Although there are identifiable reasons accounting for the large number of Genthe records in the AMICO database, the authors sought to determine whether the large number of Genthe records in the MidR Set skewed the correlations for the set. The authors isolated those 50 records and then re-ran the Pearson Product Moment test for the modified MidR Set and for the Genthe records subset. They had earlier noted the significant correlation between the artist reputation (WorldCat book records) and the number of AMICO retrievals for named artists and a stronger correlation in the case of named artists with greater than zero WorldCat records (table 3). When the Genthe records were

TABLE 4
Unique Word Counts: Means and Pearson Product Moment Correlations among AMICO Records

	HighR			MidR			MidR w/o Genthe Records		
	AMICO Records	Word Count Mean	Correlation	AMICO Records	Word Count Mean	Correlation	AMICO Records	Word Count Mean	Correlation
All records	62	59.95	0.186	363	58.20	0.357*	313	58.58	0.362*
All named artist records	60	60.80	0.185	305	57.07	0.321*	255	56.76	0.329*
Named artist records (>0 WorldCat records)	60	60.80	0.185	265	55.51	0.314*	215	55.44	0.329*
Artists with zero WorldCat records	N/A	N/A	N/A	40	67.15	0.406*	40	67.15	0.406*
All unnamed creator records	2	34.50	1.000*	58	64.23	0.502*	58	64.23	0.502*
Unnamed creators and artists with zero WorldCat records	2	34.50	1.000*	98	65.45	0.458*	98	65.45	0.458*

* Statistically significant results. Significant results for subsets in the HighR Set draw on only two records. Note: For all correlations in the LowR Set, no coefficient was returned as each record was retrieved the same number of times.

removed and the coefficients were recalculated for the same two subsets, they were found to have a slightly weaker significance. The presence of a large number of Arnold Genthe records in the sample accordingly suggests some bias. It seems, however, that the degree of bias, though measurable, is not especially large.

Putting aside the question of influence of the large number of Genthe records, it is worth noting again the varied population of records of the MidR Set. WorldCat data (number of books retrieved in a subject search in WorldCat for a named artist) exists for 265 records of the 363 comprising the MidR Set. The ranking of the WorldCat records shows a distribution close to the often observed 80/20 spread in that 80 percent of the WorldCat records are attributed to 17.4 percent of the 265 sample AMICO records whereas 20 percent of those same records account for 84 percent of the WorldCat records. By comparison, the ranking of AMICO retrievals shows a somewhat wider spread: for the 363 records in the MidR Set, 80 percent of the retrievals fall between 40 and 41 percent of the records whereas 20 percent of the records account for 58 to 59 percent of retrievals.

Comparing the use of records in the three sample sets shows the frequent occurrence of the artist name in the records of all three sets and underscores the significance of the reputation of the artist. Personally named artists and artist reputation are not the

sole factor accounting for use, however. Records without personally named artists from among the three sample sets and the wide spread of AMICO retrievals for records both with and without personally named artists point to other factors that may drive retrieval from the AMICO database.

Other Factors Contributing to Use: Record Extensiveness, Structure, Richness

A notable feature of the AMICO database is the variation in detail included in the records accompanying images. Pisciotta also remarked on the differences in the descriptions,¹³ a situation that was rather marked in the early stages of the database. The differences may be categorized as the extensiveness, structure, and richness of records.

Record Extensiveness: Word Counts

The authors considered word counts within the AMICO records as a factor contributing to the differentiation of usage among the three sets. They questioned whether more extensive descriptions account for greater frequency of use. A quantitative indicator of the extensiveness of the record is its word count. If there is a correspondence, images with more extensive records (as measured by a greater number of words) would be retrieved more than those with fewer words in the record.

The authors calculated the word counts in two ways, first by counting all the words in the record and then by counting only each unique instance of a word in a record. They designed a consistent method for counting all words. Research assistants used word-processing software in which single words, numbers, and symbols (including punctuation marks separated off by spaces) were all counted as words. Therefore, the results are relative counts using the software criteria for words rather than an absolute count of words or terms in a narrow sense.¹⁴ To get a count of unique instances of words,

the research assistants removed duplicate words, numbers, and symbols and used the software to count the number remaining in each record. Because the database search engine did not rank search results by number of occurrences of a word but, instead, retrieved an image regardless of the number of times a word appeared in the record, the authors decided that the unique word count was the more important to monitor. Therefore, the following discussion of word count refers to the count of unique instances of words in the record.

The authors speculated that if greater numbers of words contribute to higher use, the mean number of words per record in the HighR Set would be greater than those in the MidR or LowR Sets. Investigation of that hypothesis produced mixed results. (See table 4.) There were no consistent differences in records among either sets or groups of records within sets. The authors ran the Pearson Product Moment statistical test on all three sample sets as well as on subsets of the samples to better effect, however.

Among the 62 records of the HighR Set were representations of works by four different artists as well as of two works by unnamed creators. As all four artists were found with results in the WorldCat database, the category involving zero WorldCat records was not applicable (N/A). The correlation of word counts to retrievals for the works by unnamed creators was 1.000, a perfect positive correlation, likely the result of having only two such records in the set. The word counts differ out of proportion to retrieval figures. Retrievals increase as word counts increase, but the relationship between the two is not constant or exponential: the set shows no significant correlations at the 0.05 level for all records or for named creator records. The LowR Set behaved differently: all arrays returned no correlation coefficient because the number of AMICO retrievals for all records in this set was a constant (3). The results from the HighR and LowR Sets were ultimately less interesting and

useful to the authors as a result of these findings.

The MidR Set of records, by contrast to the other two, included great variation in the number of retrievals, the number of named artist records in WorldCat, and the number of words in each record, making it possibly more illustrative of the entire database. In this set, the correlation coefficients returned from the tests on word counts were significant, although not especially strong (table 4). A systematic analysis of subsets highlights the effect of word counts on usage. The correlation coefficient of word count to image retrieval for the overall MidR Set of records was 0.357. For the subset of records of works by named artists, the coefficient was lower: 0.321. Narrowing the subset down to named artists about whom the authors found records in WorldCat produced a lower (though significant) coefficient still: 0.314. This result supported the earlier conclusion that artist reputation is a significant influence in the retrieval of records in the database; the lower coefficient of word count to retrieval for named artists indicates that retrieval of these records, among the highest in the set, was driven by a factor other than word count.

Conversely, the results also point to word count as an influential factor in retrieval where the creator is unknown (or unnamed) or of minor reputation. When the subsets of named artists for whom the authors found no records in WorldCat and unnamed artists are combined, the correlation coefficient is 0.458, stronger than the coefficient for the entire MidR Set. Statistical testing on the word counts may not have yielded useful results for discussion of the HighR and LowR Sets; however, for the largest sample set, the MidR, the results were revelatory. That there is varying significance in the correlation coefficients calculated for the word counts from the MidR Set and that the significance is greatest where reputation essentially does not exist indicate that the number of words used in each record is

probably not the primary factor influencing retrieval of records from the AMICO database. Instead, word count appears to relate to retrieval of records primarily where the reputation of the work's creator is not a factor. Even though the correlation coefficients point to significance in these instances, they do not reflect great strength of correlation.

Record Structure and Richness

As indicated earlier, the record for each image is structured so that required information is included in the "creator," "work," and "ownership" headings (categories) of the public record, but other headings are available for use. Strikingly, none of the records in the HighR Set, the set with the highest percentage of artists of high reputation *and* the highest use, arranges information under more than four headings, with over 50 percent in the required three only. In both the MidR and LowR Sets, the records are constructed using three to seven of the available headings.

The authors were interested not only in categories (headings) of information provided in the records, but also in the richness (or texture) of the record, exhibited by type of terminology. By "richness," they are referring to record content, that is, informative, descriptive language that includes terms that users are likely to enter into search forms. Although the artist's name is a crucial element, records may contain names other than those of creators, such as subjects of images, owners of works, institutions, and other artists. In addition, geography and chronology are important elements. Although Wiberley's 1988 article dealt extensively with the problems of precision in defining geographic terms, he confirmed, nonetheless, that geographic terms add to the precise identification of names or other terms, as do delimiters of chronology.¹⁵ In their study, Marcia J. Bates, Deborah N. Wilde, and Susan Siegfried analyzed natural language descriptions of arts and humanities scholars' information needs and formula-

tion of search strategies.¹⁶ They identified subject searches as one category, which included geographical names (either in noun or adjectival form) and date or period (including a time modifier). The latter work shows that scholars not only use geographic and chronological terms to delimit other terms, but that there also are instances when such terms are themselves subjects of searches. Youngok Choi and Edie M. Rasmussen found that users requested images in American history most frequently by date or time period, kind of person or thing, and individual name.¹⁷ Given the findings of Wiberley, Bates et al., and Choi and Rasmussen, the authors of this study hypothesized that the presence of names, geography, and time in the records would give an indication of richness of the record that might contribute to relative use.

The authors examined selected records for the appearance of names, geographic, and date or period terms. Counting instances of this terminology presents difficulties: reconciling variations of individual or geographic/national names, deciding when a term denoting a place should count as geographic, and determining when multiple words make a single term in natural language. These difficulties accounted for the authors' decision to use a word-processing program to count single words to get a relative count of terms. These difficulties also led the authors to count only the first instance of the name, the geographic, and/or the chronological categories appearing under a heading rather than counting the number of times each appears in the records.

The authors compared a selection of records from each of the three sample sets—HighR, MidR, and LowR—for different patterns of richness. They selected records they characterized as “strong,” “countertrend,” or “remainder.”¹⁸ *Strong records* clearly fit the correlations found through statistical testing. To identify the strong records, the authors compared the highest and lowest 10 percent of the records in each sample set ranked in terms of number of retrievals and reputation or word count; records that matched high/high or low/low were considered strong. Table 5 shows the pairings for the MidR Set. The same extremes as ranked by number of retrievals and reputation or word count were used to determine *countertrend records*; however, these records showed a negative correlation between retrievals and reputation or unique word count. Of the records identified as either strong or countertrend, the authors chose approximately 10 percent of each of those combinations. *Remainder records* were identified in a stratified count of the 348 MidR Set records that remained after removing from the sets records that were counted as the extremes of strong and countertrend.

From the HighR Set, the authors identified eleven records fitting the strong criteria; out of these records the authors chose four. Similarly the authors identified and selected strong and counter trend records from the MidR Set. The authors used a somewhat different method for the LowR Set strong records because each LowR Set record had the same number of retrievals. For the LowR Set, the authors compared

TABLE 5
MidR Set, Subset Characteristics

MidR Subset	AMICO Hits/Word Counts	AMICO Hits/WorldCat Book Records
strong	high/high	high/high
strong	low/low	low/low
countertrend	low/high	low/high
countertrend	high/low	high/low

combinations of unique word counts and WorldCat book record counts rather than retrievals. They then selected one record from each of the four combinations.

To select remainder records from the MidR Set, the records were sorted in a spreadsheet using the number of OCLC WorldCat records as the first sort (highest number, zero, and not in WorldCat), number of unique words as the second sort criterion, and record identifier as the third sort criterion. Because the record identifier is an alphanumeric term based on the name of the institution that cataloged the image, there is an implicit sort by organization. From the sorted list, the authors chose every tenth record starting with the first for the remainder set. The resulting 35 records were analyzed for the appearance of name, geography, and date. Table 6 shows the results of the counts.

The comparison table shows a summary of the strong records for each of the samples and of the countertrend records and remainder records of the MidR Set, a summary of the total number of records for each set, the headings used in at least one of the records in the set, and the number and percentage of records with at least one appearance of individual names, geographic terms, and date or period terms. From these data, the authors constructed an index of richness of records within each set. It consisted of summing the number of name, geographic term, and date or period appearances (that is, that the type of term appeared at least once) within each heading (creator, work, ownership, etc.) and dividing it by the number of records (whether or not the heading is used in the record). For example, in the strong HighR subset, there are four records. Under the creator heading, there are three records with the appearance of at least one name, three records with the appearance of at least one geographic term, and three records with the appearance of at least one chronological term for a total of 9 out of a possible 12 (four records with the possibility of the three elements).

Among the strong records in the three subsets (14 records), name, geography, and date or time period all appeared. Not unexpectedly, the creator and ownership headings were most rich, whereas the title of work was midrange. Among the required fields, the lowest indication of richness was in the title of work category of the strong LowR records, at 40 percent. The countertrend and remainder records of the MidR Set also follow a similar pattern with the highest percentage of appearances among the required categories of creator and ownership and slightly lower percentages in the title heading. With the exception of the commentary heading in the HighR Set, the optional headings are used less frequently in all groups of records, showing an appearance rate of less than 45 percent for name, geographical, and date terminology.

To summarize: the appearances of name, geography, and date terms show no strongly contrasting patterns among the sets. On the other hand, the percentages confirm earlier findings. The name element has the lowest percentage of appearances in the strong LowR Set records, consistent with the trend of high use associated with artist name reputation. Date terms are the second lowest percentage in the same set. The name element has the lowest percentage among the remainder MidR Set records. This suggests that a consideration for further study is whether images of works that are not well known have greater chances of usage through extensive optional name, geographical, and date-related information in the record.

Record Richness: Forms of Terminology

The analysis so far has noted the presence of each of three categories—names, geographical terms, or date terms—in all the sample records without regard to forms of terminology; furthermore, it did not note how often the categorized terms appeared as unique or repeated instances. The following analysis of a limited number of records examines the forms in which the

TABLE 6
Records with at Least One Appearance of Name, Geographic Term, and Date by Heading in Selected Record Subsets

<i>Heading</i>	Strong HighR records (n = 4)				Strong LowR records (n = 5)				% of possible instances
	Name, records (% of records)	Geography, records (% of records)	Date, records (% of records)	% of possible instances	Name, records (% of records)	Geography, records (% of records)	Date, records (% of records)	% of possible instances	
Creator	3 (75.00)	3 (75.00)	3 (75.00)	66.67 ^a	3 (60.00)	5 (100.00)	3 (60.00)	73.33 ^b	
Work	3 (75.00)	1 (25.00)	2 (50.00)	50.00	1 (20.00)	2 (40.00)	3 (60.00)	40.00	
Ownership	4 (100.00)	4 (100.00)	2 (50.00)	83.33	2 (40.00)	5 (100.00)	3 (60.00)	66.67	
Commentary	2 (50.00)	2 (50.00)	2 (50.00)	50.00	0	1 (20.00)	0	6.67	
	Strong MidR records (n = 5)				Countertrend MidR records (n = 10)				
Creator	4 (80.00)	4 (80.00)	3 (60.00)	73.33 ^c	8 (80.00)	7 (70.00)	7 (70.00)	73.33 ^d	
Work	1 (20.00)	4 (80.00)	4 (80.00)	60.00	0	6 (60.00)	10 (100.00)	53.33	
Ownership	3 (60.00)	5 (100.00)	2 (40.00)	66.67	8 (80.00)	10 (100.00)	4 (40.00)	73.33	
History					3 (30.00)	3 (30.00)	3 (30.00)	30.00	
Commentary	1 (20.00)	0	1 (20.00)	13.33	4 (40.00)	4 (40.00)	5 (50.00)	43.33	
Related Materials					2 (20.00)	1 (10.00)	2 (20.00)	16.67	
	Remainder MidR records (n = 35)								
<i>Heading</i>	Name, records (% of records)	Geography, records (% of records)	Date, records (% of records)	% of possible instances	Name, records (% of records)	Geography, records (% of records)	Date, records (% of records)	% of possible instances	
Creator	30 (85.71)	27 (77.14)	27 (77.14)	73.33 ^e	20 (57.14)	20 (57.14)	20 (57.14)	73.33 ^e	
Work	16 (45.71)	19 (54.29)	19 (54.29)	60.95	29 (82.86)	29 (82.86)	29 (82.86)	60.95	

TABLE 6
Records with at Least One Appearance of Name, Geographic Term, and Date by Heading in Selected Record Subsets

Heading	Name, records (% of records)	Geography, records (% of records)	Date, records (% of records)	% of possible instances
Ownership	24 (68.57)	35 (100.00)	7 (20.00)	62.86
History	5 (14.29)	0	5 (14.29)	9.52
Commentary	2 (5.71)	4 (11.43)	3 (8.57)	8.57
Related Materials	5 (14.29)	1 (2.86)	0	5.71

^aFor all strong HighR records, 3 categories times 4 records = possible 12 category appearances (instances)
^bFor all strong LowR records, 3 categories times 5 records = possible 15 category appearances (instances)
^cFor all strong MidR records, 3 categories times 5 records = possible 15 category appearances (instances)
^dFor all countertrend MidR records, 3 categories times 10 records = possible 30 category appearances (instances)
^eFor all remainder MidR records, 3 categories times 35 records = possible 105 category appearances (instances)

categories appear in the records. The aim of this exercise is to suggest multiple ways of constructing records that incorporate elements associated with high use. Because of the similarity in the records, as shown below, the beginning analysis of two records establishes a pattern whereas comments on an additional seven records note only characteristics that differ from that analysis.

The authors began with two contrasting records from the strong MidR records subset. The two selected images fit the expected trend among the MidR records; that is, higher usage generally is found for image records for which the artists have a higher reputation or, alternatively, when a record of a work by an unnamed creator or an artist of no reputation has a higher number of unique words. The records were for an image from Duane Michals's photographic series *Paradise Regained*¹⁹ and an image of Jan Martss's print *Dutch Cavalrymen in Action*.²⁰ In this analysis, the focus is on the primary categories of individual names, geographical terms, and date-related terms in each record without regard to the headings within the record. Although there are many points of contrast between the two records, the appearances of names, geographical terms, and dates show a great deal of similarity.

In both records, the forms of names vary among given names and surnames as well as surnames alone. The names identify individuals such as an artist (e.g., Jan Martss) and organizations or legal entities such as Achenbach Foundation for Graphic Arts. Geographical terms include city, state, or country, alone or in combinations such as San Francisco, California, USA. In addition, a geographical term is used as part of the official name of an organization, for example, the Cleveland Museum of Art. Two forms of date appear in these records, as year and as century, to locate the person or object in time. The number of appearances of each category varies by the type of information included in the catalog record.

For example, the record describing the image of Michals's photograph includes extensive biographical information on Michals, naming other artists as well. Recounting Michals's travels or exhibits of his works necessitated many unique and duplicated geographical terms. The cataloger's analysis of the photograph contains names that might be categorized as myth, religion, or literature.

As a potential contrast to the two records exhibiting expected trends in the relation between use and reputation or unique word count, the authors examined two records from among those illustrating the trend counter to an expectation of factors influencing usage. The first of these two is a sculpture entitled *Kero*, attributed to an unnamed artist of the Inca culture.²¹ The other is a photograph, *Coast View*, by Edward Weston.²²

The contrast between this pair of records and the previous regarding the use of name, geographical terms, and dates is less strong than between the two records of this pair. Little information is available in the record for the Inca sculpture. The only name is that of the donor, using given name, middle initial, and surname. Only one example of a geographic term appears in the description: city, state, and country are part of the owning institution's name. A date appears just once, as years in a range, approximating the sculpture's age.

The second record of this pair describes the photographic print by Edward Weston, *Coast View*. As noted, it is more extensive than the *Kero* record. The artist's given and surnames appear as creator, in titles of books listed in the record, and as a named collection in the museum. Other book authors' names appear as surnames. Geographical terms appear as locations; as city, state, and country, together or alone, in varying sequences; and as part of an organization's name. An adjectival form of a country's name is used to provide identifying information about the artist. Publishing dates are shown as years. Unusual for any of the records the

authors examined in detail, this Weston record documents precise dates such as birth and death, appearing as month, date, year; years also appear in ranges. This record, then, fits the original pattern set by the two strong MidR sets.

The analysis of records in the authors' original sample MidR Set so far has looked at two records exhibiting the expected trends regarding the relation between retrieval and reputation or high unique word count; it also has looked at records that appear counter to the trend. This analysis concludes with the remainder records; there were thirty-five such records. The authors drew five records from the remainder MidR subset in which to examine the presence of name, geography, and date terms. (See table 7.) Each record was contributed by a different institution and ranged in number of records retrieved in WorldCat and number of AMICO retrievals.

The first record, *Stigmatization of St. Francis*, is an example of a record without a named artist.²³ The second record is the Henry Moore image *Maquette for Head*.²⁴ The third record describes a photograph of Mrs. Joseph B. Chamberlain by Arnold Genthe, one of the great number of images done by Genthe appearing in the authors' original sample.²⁵ The fourth record, *Fan*, is an example of a record describing an image by a named creator (George Keiswetter) for whom there were no records in WorldCat.²⁶ The fifth record describes the Picasso image *Femme Torero II*.²⁷

The analysis by name, geography, and date is similar to the analyses of the other two sets with some variation. The *Stigmatization* record, for example, has no artist named, but names appear as the subject of the work or in the extensive commentary that tells the story behind the religious subject of the diptych. Most names are of Christian saints but also include the names of an aristocrat thought to be the work's first owner and her husband. The forms of names include given names and surnames; an adjectival form of a name

TABLE 7
Detailed Examination of Select MidR Records

	Record 1	Record 2	Record 3	Record 4	Record 5
Creator	Henry Moore	George Keiswetter	Pablo Picasso	Unknown (Italian)	Arnold Genthe
Title	Maquette for Head: Lines	Fan	Femme Torero II	Stigmatization of St. Francis	Chamberlain, Joseph B., Mrs., with dog
AMICO hits	11	9	306	83	9
WorldCat records	665	0	3,665	No searchable artist name	14
Word count	24	63	53	138	61
<i>Headings</i>					
Creator	N, G	N, G, D	N, G, D	G	N, D
Work	X	G, D	G, D	N, D	N, D
Ownership	N, G, D	G	N, G	N, G	G
History	N/A	N/A	N/A	N/A	N, D
Commentary	N/A	N/A	N/A	N, G, D	G
Related Materials	N/A	N/A	N/A	N/A	N
Descriptive term	X	N/A	X	N/A	X
<i>Note:</i> N = at least one appearance of a name; G = at least one appearance of a geography term; D = at least one appearance of a date; X = heading is used in the record description, but no appearance of name, geography, or date terms; N/A = heading not used in the record description.					

appears. Geographical terms include the adjectival form of the country to help identify the origin of the work, a city, state, and country combination, as well as city alone. Dates appear in decade and century forms, unsurprising for a work several hundred years old.

The Moore record differs in that there is one name, Moore's own as artist and donor, in the record. The only date appears as the year of the donation. Genthe's record includes features that distinguish it from many of the other records analyzed. No country of origin for the artist is noted, although other geographic terms appear, and, interestingly, a street address is given as the context. Dates include the month, day, and year to describe the work as well as other years. Keiswetter's *Fan* is a bejeweled costume piece painted for a fan company. The forms of names in

the record include a surname as part of a company name and a surname as part of a named collection. The final record, Pablo Picasso's *Femme Torero II*, follows the pattern of records in the remainder MidR subset as well as in the strong and countertrend MidR subsets. Use of name, geography, and date in this record with an artist of high reputation as measured by WorldCat records is remarkable only in its conformity to a pattern found in many other records.

Between them, Wiberley and Bates et al. have noted the frequent occurrence of names, geography, and dates in the work of humanists as indexing terms and as the kind of search terms used by scholars. In the records of the AMICO image database, the catalog records follow suit. Regardless of what other information is available, most records include all three

kinds of vocabulary in some form or other. Identification of one individual among many is important in Western cultures. Situating a person or an item in space and time aids identification and helps set the context of an artist or a work in order to understand it better. This analysis has indicated no striking contrast in the use of name, date, and geography that might distinguish high-use records from low-use records. It has supported the analyses of Bates et al. and Wiberley regarding the prevalence of names, dates, and geography.

Although catalog records for images have similar terminology to both entry terms in reference works or indexing services and to the language of scholars defining their information needs or research questions, they have a potential for abundant description utilizing common terms. Keyword searching of the AMICO database will identify the record if the terms appear in any part of the record. Both Wiberley and Bates et al. include common terms in their analyses. Both also note the difficulty in identifying and classifying such terms.²⁸ The authors of this study experienced similar difficulty. Thus, the most productive approach they could take was to use the unique word count as a surrogate for the extensiveness of a record and the instance of a name, geographic term, or chronological term as an indicator of the richness of the record. Further exploration of common terms in the description of art images could be undertaken in future studies.

Conclusions

Investigation into use statistics for the AMICO database points to implications about the structure of records in an image database that could influence the description of objects to be digitized. Moreover, the study suggests patterns for further investigation with implications for collaborative digitization endeavors.

Among the factors studied, creator reputation drives retrieval of images in the sample. The prevalence of a few widely

recognized artist names in the HighR Set of most-retrieved records suggested as much. Statistical testing subsequently established that the trend was significant. This finding is somewhat weakened by the presence of one artist's name in both the HighR and MidR sample sets and of fifteen names in both the MidR and LowR sample sets. Because the usage samples are based on images and not on creators, it is no surprise that users may be more interested in one or another particular image by the artist. Further muddying the waters is that the sample data sets are not tied to use strategies, and the authors do not have specific knowledge whether users have searched for artists' names, and if so, whether they have searched as keyword or by creator field.

One further limitation arises from the method of using the number of monographic records in WorldCat about an artist as surrogate for the artist's reputation, at least for some art forms as noted earlier. Anne McCauley's study of photography's history documents what she calls the "near invisibility" of the photographer as an artist,²⁹ suggesting the need for additional comparative indications of reputation among artists. Nonetheless, the correlation of the artist's reputation to usage is consistent with other research showing the importance of the artist-creator's name and reputation.

Decisions about which works to digitize may not necessarily depend on an artist's reputation. Instead, institutions may prioritize works by more obscure creators to meet local needs or to illustrate local collection strengths. In these cases, the research suggests that extensive descriptions have a positive influence on retrieval of the records. The authors found correlation simply between the number of unique instances of words used in the image record and the number of retrievals. The correlation was stronger in cases where a creator was unknown or the image was not attributed to a named person.

The authors looked further than simple word counts, examining what they termed

richness of language. Building on previous research about how humanities scholars search databases and on research analyzing the terminology of entry terms in encyclopedias, dictionaries, and indexing services, the authors found potentially significant trends in the inclusion of language of personal names, geographic location names, and dates: all three can improve the retrieval of images. It is difficult to establish what constitutes a "term" and how the form of the term affects retrieval, however. Still, to improve the likelihood of retrieval of images, writers of records are advised to build richness into the record by including the types of terms noted earlier; if such terms are few, writers are advised to include at least a variety of terms. A search engine that permits truncation will further increase chances of retrieval when there are multiple forms of significant terms.

This study has implications for constructing an image database, either by a single agency or in collaboration with others. Institutions building image databases

could develop tiers of standardization. They would take into account the reputation of the artist with requirements for extensive and rich description for works with no creator or with one who is less known. Further studies are needed to confirm the findings of this one.

There were limits on what the record samples could represent. As a result, some factors remain unknown. Comparing records by contributing institutions might indicate how practice in applying the AMICO cataloging standards varies and whether usage varies by institution. Comparing records of multiple works of a single artist could help confirm or discount the correlation of word counts to retrievals. Finally, studies of transaction logs and of database users' needs could provide information on how users navigate the Web site and for what reasons (scholarship, school assignment, interest). These studies could identify factors to include in constructing an image database.

Notes

1. Youngok Choi and Edie M. Rasmussen, "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, no. 6 (2003): 498–511.

2. Deborah D. Blecic, Joan B. Fiscella, and Stephen E. Wiberley Jr., "The Measurement of Use of Web-based Information Resources: An Early Look at Vendor-supplied Data," *College and Research Libraries* 62, no. 5 (2001): 434–53.

3. Robert V. Krejcie and Daryle W. Morgan, "Determining Sample Size for Research Activities," *Educational and Psychological Measurement* 30, no. 3 (1970): 608.

4. Random numbers were drawn using Geoffrey C. Urbaniak and Scott Plous, *Research Randomizer*. Available online from <http://www.randomizer.org/form.htm>. [Accessed 7 March 2002].

5. By "usable," the authors mean those records that were researchable in the AMICO database. The usage data from the database included record identifiers as provided by the contributing institution. For some institutions, the identifiers were searchable, but for others, the authors would have been forced to guess at matches. The authors were able to contract an RLG employee to map the identifiers from the usage data to identifiers that were searchable in the AMICO user interface. A small number of these, however, still yielded no results.

6. Stephen E. Wiberley Jr., "Subject Access in the Humanities and the Precision of the Humanist's Vocabulary," *Library Quarterly* 53, no. 4 (1983): 420–23.

7. ———, "Names in Space and Time: The Indexing Vocabulary of the Humanities," *Library Quarterly* 58, no. 1 (1988): 24–25.

8. Linda H. Armitage and Peter G. B. Enser, "Analysis of User Need in Image Archives," *Journal of Information Science* 23, no. 4 (1997): 288.

9. Examples of personally named artists from the record set include the famous (Pablo Picasso, Claude Monet) and the seemingly less famous (Arnold Genthe, Alice Barber Stephens). Examples of other named creators from the record set include cultures and geographic locales (Inca, Eastern Anatolia) and studios, factories, and artisans identified only by an association (Sèvres Factory, Court atelier of Duke Ludwig I).

10. In his work, "A Methodological Approach to Developing Bibliometric Models of Types of Humanities Scholarship" (*Library Quarterly* 73, no. 2 (2003): 121–59), Wiberley uses the number of author entries in a library catalog as a surrogate for the amount of publishing done by that author. The amount of publishing is an indicator of the author's prominence.

11. Henry Pisciotta, "Art Museum Image Consortium: The AMICO Library," *CAA Reviews*, May 25, 2000. Available online from <http://kola.forest.net/colleageart/action.lasso>. [Accessed 25 March 2005].

12. Renata V. Shaw, compiler., *A Century of Photographs 1846–1946: Selected from the Collection of the Library of Congress* (Washington D.C.: Library of Congress, 1980).

13. Pisciotta, "Art Museum Image Consortium."

14. The AMICO database defaults to keyword searching. Testing of the keyword search function showed that common stop words, abbreviations, and even Boolean operators were searched as if they were keywords as well. The resulting complexity in writing rules for the treatment of stop words and similar terms in conducting the word counts contributed to the decision to use a word-processing program's word count tool.

15. Wiberley, "Names in Space and Time."

16. Marcia J. Bates, Deborah N. Wilde, and Susan Siegfried, "An Analysis of Search Terminology Used by Humanities Scholars: The Getty Online Searching Project Report Number 1," *Library Quarterly* 63, no. 1 (1993): 1–39.

17. Choi and Rasmussen, "Searching for Images."

18. To determine which records were "strong" and which were "countertrend," the authors found the top 10 percent and bottom 10 percent of the MidR records when sorted by AMICO retrievals, OCLC WorldCat record counts, and unique words. For a record to be considered strong, it had to rank in the top 10 percent for both AMICO retrievals and OCLC records. It also could rank in the bottom 10 percent of both. A third possibility involved a record ranking in the bottom 10 percent of OCLC records, ranking in the top 10 percent of AMICO retrievals, and ranking in the top 10 percent of unique word counts. The strong records acted in a manner predicted by earlier statistical work. For a record to be considered countertrend, it had to rank in the top 10 percent of either AMICO retrievals or OCLC records but rank in the bottom 10 percent of the other. In the case where OCLC record counts were low and AMICO retrievals were high, if the word count was not high as expected, the record was considered to be countertrend. In other words, these records acted in a manner counter to predictions based on earlier statistical work.

19. Duane Michals, *Paradise Regained (Number 30)*, 1968. © The Cleveland Museum of Art, Cleveland OH, The AMICO Library: CMA_.1989.446.a.

20. Jan Martss, *Dutch Cavalrymen in Action, 17th century*, The Fine Arts Museums of San Francisco, San Francisco, CA, 1963.30.10683, The AMICO Library: FASF.1228.

21. [Inca], *Kero, 1300-1550*, The Minneapolis Institute of Arts, Minneapolis, MN, 98.163.2, The AMICO Library: MIA_.98.163.2.

22. Edward Weston, *Coast View, Point Lobos, USA*, 1938, Center for Creative Photography, Tucson, AZ, 81:251:209, The AMICO Library: CCP_.81:251:209.

23. [Unknown (Italian)], *Stigmatization of St. Francis*, circa 1330, J. Paul Getty Museum, Los Angeles, CA, 86.PB.490, The AMICO Library: JPGM.86.PB.490.

24. Henry Moore, *Maquette for Head: Lines*, The Art Gallery of Ontario, Toronto, ON, 74/55, The AMICO Library: AGO_.74/55.

25. Arnold Genthe, *Chamberlain, Joseph B., Mrs., with dog*, October 23, 1917, Library of Congress Prints and Photographs Division, Washington, DC, LC-G432-2072, The AMICO Library: LOC_agc96006659/PP.

26. George Keiswetter, *Fan*, circa 1890, The Museum of Fine Arts, Boston, Boston, MA, 1976.369, The AMICO Library: BMFA.1976.369.

27. Pablo Picasso, *Femme Torero II (Female bullfighter)*, 1934, The Fine Arts Museums of San Francisco, San Francisco, CA, 1971.28.72. © Estate of Pablo Picasso/Artists Rights Society (ARS), New York, The AMICO Library: FASF.182.

28. Wiberley, "Names in Space and Time," 4, 5; Bates, Wilde, and Siegfried, "An Analysis of Search Terminology Used by Humanities Scholars," 8–11.

29. Anne McCauley. "Writing Photography's History before Newhall," *History of Photography*, 21, no. 2 (summer 1997): 87–101.