

PEASE AWARD

Providing Subject Access to Images: A Study of User Queries

Karen Collins

Abstract

This paper describes a study of user queries conducted at two historical photographic collections—the North Carolina Collection at the University of North Carolina at Chapel Hill, and the North Carolina State Archives in Raleigh. Patron requests were analyzed in order to determine which types of subject terms and attributes of images are used most often in requests for photographs. Basic categories of terms were created, and the number of requests utilizing each category of term was tallied. It was found that subject terms, both generic and specific, were used far more frequently than any other categories of terms in requests for photographs. Generic subject terms appeared most often in requests, indicating the importance of these terms for indexing. Time and place were the next most commonly used types of terms. In contrast, genre, visual terms, format, and creator/provenance were mentioned relatively infrequently.

Introduction

Suppose you wish to locate an image of Santa Barbara, California, circa 1925. How should you go about searching for such an image? In the words of Todd Ellison, the best approach is to “think like an archivist.”¹

¹ Todd Ellison, “Access to Photographs,” *Colorado Libraries* 20 (Summer 1994): 44.

Karen Collins was awarded the 1997 Theodore Calvin Pease Award for this paper, which was written for a course taught by Helen Tibbo in the School of Information and Library Science at the University of North Carolina at Chapel Hill. The Pease Award is named for the first editor of the American Archivist and is given to the best student paper as judged by the current editor of the American Archivist and two individuals with expertise in archival research and literature. The 1997 selection committee consisted of Philip Eppard, Megan Sniffin-Marinoff, and Elizabeth Yakel. The award was presented on August 28, 1997, at the SAA annual meeting in Chicago.

In other words, try to figure out what organization or person might have created such an image, then determine where that person or organization's photographs are held. For example, you might reason that Santa Barbara, having been a tourist destination in the 1920s, would be the subject of postcards. In fact, the Detroit Publishing Company, a major publisher of postcards during that era, did photograph Santa Barbara. This agency's photograph collection is held by the Library of Congress' Prints and Photographs Division. Contacting the Library of Congress is likely to result in the location of several images matching your criteria.

Unfortunately, the above example is extremely contrived. Suppose you don't know about the Detroit Publishing Company, or it doesn't occur to you to research postcard manufacturers. Furthermore, you may have no idea where the photographs of the Detroit Publishing Company (now defunct) are currently held. This is a far more likely scenario, and one that begins to point to some of the difficulties in locating visual information.

Because it is common and useful to label photographs according to geographic location and chronology, the above request is simpler than most. In a study of user queries at the Prints and Photographs Collection of the National Library of Medicine, it was estimated that approximately one-third to one-half of all requests are "image construct queries," i.e., requests describing the desired image in terms of its visual components.² Fulfilling this type of request requires a far more detailed item-level description than that usually given to archival collections.

An additional difficulty lies in the fact that the same image may mean different things to different people, and it may be needed for many different purposes. Sociologists, historians, scholars of art and architecture, graphic designers, picture researchers, educators, and others all use images in different ways. Given the increasingly interdisciplinary nature of research, it is desirable that a collection of images be searchable by persons from any field and for a great variety of purposes. In many institutions, however, retrieving an image requires knowledge of its creator or of its title, supplied by either the image creator or the cataloger.³ Often, only collection-level descriptions are available. This means that image construct queries are simply not possible, and the patron must either have specialized subject knowledge or rely heavily on the memories of individual collection managers.

Because visual information is complex and not exactly translatable into verbal language, describing the subject matter of images is uniquely problem-

² Lucinda H. Keister, "User Types and Queries: Impact on Image Access Systems," in *Challenges in Indexing Electronic Text and Images*, edited by Raya Fidel (Medford, N.J.: Learned Information for the American Society of Information Science, 1994), 13.

³ Karen Markey, *Subject Access to Visual Resources Collections: A Model for Computer Construction of Thematic Catalogs* (New York: Greenwood Press, 1986), 3.

atic.⁴ As a result, it is usually necessary for a patron to visually examine several images in order to decide on the best one for a given purpose. Systems that allow for the scanning of thumbnail images or other visual surrogates are therefore being developed. Progress is also being made toward creating systems that use images ("picture queries") to retrieve other images.⁵ In most cases, however, it is still necessary to use some form of verbal request in order to retrieve a subset of images for examination. Therefore, research into ways to more effectively index visual materials remains important.

Several theoretical models, thesauri, and classification systems have been developed to aid in the indexing and retrieval of images. Although all of these methods contain assumptions regarding the visual information needs of patrons, very little research has been done to determine what these needs actually are. Some issues that need to be examined by such research include: 1) whether there is a need for more indexing at the pre-iconographical level (defined below), 2) whether there are attributes of images that are not being indexed but would be useful in retrieving images, and 3) whether some attributes are more important to index than others. This study attempts to shed light on these questions by analyzing patron requests for photographs in two photographic archives.

Literature Review

Theoretical Models. The philosophical work of the late art historian Erwin Panofsky, as interpreted by Sara Shatford, Karen Markey, and others, has been influential on many people working to develop systems for subject access to images.⁶ Panofsky recognized three successive levels of analysis in the study of visual materials, which he termed *pre-iconographical description*, *iconographical analysis*, and *iconological interpretation*.⁷ Pre-iconographical description corresponds to the identification of forms as representations of objects and events, and it requires only the knowledge gained from everyday experience. This "primary subject matter" can be either *factual* or *expressional*. For example, a person might recognize that an image represents a house (factual subject matter) and also that it conveys a cozy or peaceful mood (expressional subject matter). Although the expressional qualities of an image are relatively

⁴ Elaine Svenonius, "Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages," *Journal of the American Society for Information Science* 45 (Sept. 1994): 600-6.

⁵ A.E. Cawkell, "Picture-queries and Picture Databases," *Journal of Information Science* 19 (1993): 409-23.

⁶ Sara Shatford, "Analyzing the Subject of a Picture: A Theoretical Approach," *Cataloging & Classification Quarterly* 6 (Spring 1986): 39-61; Markey, *Subject Access to Visual Resources Collections*; and Karen Markey, "Access to Iconographical Research Collections," *Library Trends* 37 (Fall 1988): 154-74.

⁷ Markey, *Subject Access to Visual Resources Collections*, 1.

subjective, they are still available to those with only “everyday familiarity with objects and events.”⁸

Iconographical analysis, or the identification of “secondary subject matter,” corresponds to a higher level of interpretation of an image. It requires the viewer to have not only everyday experience but also “knowledge of literary sources, customs, and cultural traditions peculiar to a certain civilization.”⁹ For example, a photograph of a man in military uniform holding pieces of what appear to be aluminum foil takes on added significance when one knows that the person is Major Jesse Marcel, and that the material in his hands has been conjectured to be the remains of a crashed flying saucer. This knowledge is built upon the factual and expressional meaning of the picture but also requires familiarity with the 1947 Roswell crash and its surrounding lore.

Panofsky’s third level of analysis, iconological interpretation, corresponds to “the identification of underlying principles which reveal the basic attitude of a nation, period, religion, class, or philosophical persuasion.”¹⁰ This level of meaning is highly interpretive, and cannot be agreed upon with any consistency. Although indexing or cataloging images at this level is therefore usually not practical outside of the field of art, it is important to keep in mind that valid iconological interpretation relies upon “accurate pre-iconographic description and correct iconographical analysis of a picture.”¹¹

Much of the literature about subject cataloging and indexing of images emphasizes the need for greater access at the pre-iconographical or primary subject matter level of description.¹² The information contained in a catalog record generally describes an image’s secondary subject matter (e.g., “Maj. Jesse Marcel holding debris from Roswell crash”). This means that a patron must have a certain amount of specialized knowledge to find an image, and that, in general, one cannot search across a collection for all images of people holding debris, or of men in uniform, or of any other generic class of depictions. It also means that the cataloger or indexer must go through two levels of interpretation to arrive at a description, resulting in a greater likelihood of discrepancy from one catalog to another. As Karen Markey points out, “Anyone searching a collection of visual images accessible [only] by second-

⁸ Shatford, “Analyzing the Subject of a Picture,” 43.

⁹ Markey, “Access to Iconographical Research Collections,” 156.

¹⁰ Erwin Panofsky, quoted in Markey, “Access to Iconographical Research Collections,” 156.

¹¹ Shatford, “Analyzing the Subject of a Picture,” 45.

¹² See, e.g., Jeanne M. Keefe, “The Image as Document: Descriptive Programs at Rensselaer,” *Library Trends* 38 (Spring 1990): 659-81; Keister, “User Types and Queries,” 7-19; Chi Hong C. Leung, D. Hibler, and N. Mwara, “Picture Retrieval by Content Description,” *Journal of Information Science* 18 (1992): 111-19; Markey, *Subject Access to Visual Resources Collections*; Markey, “Access to Iconographical Research Collections,” 154-74; and Shatford, “Analyzing the Subject of a Picture,” 39-61.

ary subject matter is at the mercy of the indexer's interpretation."¹³ Describing images by their primary subject matter preserves information that both specialists and nonspecialists can use to gain access to the collection.

Patrons of image collections come from many different fields and professions, with needs ranging from the study of techniques or styles to the obtaining of an illustration for a lecture or book. Among those who use picture collections are publishers, historians, social scientists, graphic designers, museum professionals, architects, genealogists, teachers, students, artists, writers, and others. Given this interdisciplinary need for images, it is likely that many potential users would benefit from a more detailed description of the subject matter of pictorial collections. Jeanne Keefe gives an example of how added subject headings for holdings of the Rensselaer Architecture Library's Slide Collection provided greater access for a wider variety of patrons:

While the title (Sydney Opera House) is still the primary notation, the slide document itself also contains information on a variety of subjects (e.g., Ridge Beams, Glass Curtain Walls, Tiles, Shell Vaults, Precast Concrete Ribs, Concert Halls, etc.). These different references now make that slide available to those patrons needing examples of different types of materials, structures, and/or designs. It means that a slide of a statue in a fifteenth-century Gothic cathedral is now available to the student or professor of Medieval history who needs examples of armor or dress from the Middle Ages. Viewing a slide as a document instead of a composition significantly increases its usefulness as a visual resource.¹⁴

Jackie Dooley argues that as research has become increasingly interdisciplinary, and as archival materials are routinely included in integrated online catalogs, direct subject access to these materials is becoming more necessary.¹⁵ Although she is speaking specifically of archival materials, her statements apply also to image collections, whether or not they are archival. She asks, "Is it possible that materials currently accessible only through provenance might find their way to additional users if a variety of subject access points were added?"¹⁶ Such considerations are important in these times when libraries and archives are frequently being asked to justify their existence.

Karen Markey describes two examples of image repositories that have attempted to provide access to their holdings by primary subject matter. The Repository of Stolen Art was developed by the Royal Canadian Mounted Police for the identification and recovery of stolen cultural property, including "objects from soils or waters, ethnographic art, military objects, books, rec-

¹³ Markey, *Subject Access to Visual Resources Collections*, 7.

¹⁴ Keefe, "The Image as Document," 663.

¹⁵ Jackie M. Dooley, "Subject Indexing in Context," *American Archivist* 55 (Spring 1992): 345.

¹⁶ Dooley, "Subject Indexing in Context," 351.

ords, photographs, and sound recordings.”¹⁷ Users who do not have any sort of specialized training in art are expected to be able to describe works using this system. The Historic New Orleans Collection also attempts to provide access to both specialist and nonspecialist users by including both primary and secondary subject matter descriptions of its approximately 150,000 photographs, prints, and paintings.¹⁸ These efforts should help show whether this depth of indexing is possible and of benefit to patrons.

Sara Shatford expands upon the first two levels of Panofsky’s model to include the distinction between what an image is *of* (objects, creatures, and events), and what it is *about* (“symbolic meanings and abstract concepts that are communicated by images in the picture”).¹⁹ She points out that although indexing “aboutness” is relatively subjective, it can be worthwhile to do so for some collections. Krause also argues that a greater amount of such “soft indexing” could improve access for some patrons.²⁰ Although “aboutness” and the emotional or expressional qualities of an image are difficult to agree upon, it is often precisely these subjective qualities that a patron is looking for, as in Keister’s example of a patron searching for a “warm picture of a nurse, mother, and baby” for publication in a public health journal.²¹ Shatford asks:

Rather than pretend that certain aspects of picture indexing are objective, would it not be better to admit that subjectivity exists, acknowledge its drawbacks (chiefly, that it leads to inconsistency) and recognize that subjective judgments and analysis can provide valuable access to information?²²

Shatford additionally makes the important observation that images are always both generic and specific, and that one or both aspects can be indexed. Her classification of the possible subjects of an image is based upon these distinctions, combined with the basic subject facets of *Who? What? When? and Where?*²³

In a later paper, Shatford Layne presents a list of attributes that would ideally be included in the catalog description of an image:

- *biographical* attributes: including creator, ownership record, location, and value history.

¹⁷ Markey, “Access to Iconographical Research Collections,” 167.

¹⁸ Markey, “Access to Iconographical Research Collections,” 167.

¹⁹ Shatford, “Analyzing the Subject of a Picture,” 47.

²⁰ Michael G. Krause, “Intellectual Problems of Indexing Picture Collections,” *Audiovisual Librarian* 14 (May 1988): 73–81.

²¹ Keister, “User Types and Queries,” 10.

²² Shatford, “Analyzing the Subject of a Picture,” 57.

²³ Shatford, “Analyzing the Subject of a Picture,” 49.

- *subject*: classifiable into four facets: time, space, activities/events, and objects.
- *exemplified* attributes: the physical form of an image.
- *relationship* attributes: the linkage of an image with other images/texts (e.g., preliminary sketch and final painting).²⁴

While this list presents a rather complete set of characteristics by which people might describe photographs, it remains to be seen whether some attributes might be more important than others to patrons seeking visual information.

Visual Access to Visual Resources. No matter how many words are used to describe an image, the *visual* information contained will never be completely captured. Elaine Svenonius argues that the languages of images and music cannot be fully translated into words: "What is expressed cannot be spoken of; it cannot be referred to using language; it cannot be named and cannot be indexed by index terms."²⁵ Hence the importance of having an opportunity to examine the image itself, or an image of the image. As Lucinda Keister states,

Presentation of an image surrogate...addresses the aesthetic or emotional need of the user—a highly subjective need not appropriate for the cataloger to consider. Any picture reference librarian knows that patrons ask for "dramatic pictures," "grabbers," etc. . . . Watching patrons searching images...shows that a most interesting dynamic occurs in which words, in carefully constructed catalog records, introduce the user to selections of images; and the user then reviews, analyzes, and verifies with words again before finally arriving at the selection. The user constantly checks...to see which image "works," that is, communicates most effectively the desired message.²⁶

Although specifying the subject, time period, geographical region, format, etc. of an image can narrow a search considerably, the patron typically must visually examine many images to determine which ones meet the subjective needs described above. As Besser notes, "This creates work for the library/repository (which must retrieve many unneeded delicate images), wear and tear on the collection, and a great deal of inconvenience for the user."²⁷ Image surrogates, in the form of photocopies or digital thumbnail images, provide a solution to this problem.

²⁴ Sara Shatford Layne, "Some Issues in the Indexing of Images," *Journal of the American Society for Information Science* 45 (Sept. 1994): 585–88.

²⁵ Svenonius, "Access to Nonbook Materials," 603.

²⁶ Keister, "User Types and Queries," 17.

²⁷ Howard Besser, "Visual Access to Visual Images: The UC Berkeley Image Database Project," *Library Trends* 38 (Spring 1990): 787.

Digital image databases have been developed with the capability of presenting several thumbnail images for examination at one time. A.E. Cawkell points out that such systems "immediately bring to bear the most efficient selective system by far—the human eye/brain."²⁸ The user can quickly scan a number of images and select the ones of interest for enlargement or additional information.

Because verbal indexing of images is problematic, systems are being developed to allow for the retrieval of images using other images. Cawkell describes several such prototypes being used with small test collections.²⁹ With these systems the user presents a query picture consisting of either a rough sketch or, in some cases, another image. Various methods are then used to retrieve images containing similar visual features.

Although systems using query pictures hold promise for improving retrieval with certain types of queries, e.g., "photographs of automobiles," these represent only a limited subset of patron requests. Because much identifying information (e.g., date, geographical location, event, etc.) is not to be found in the image itself, it is highly unlikely that query pictures alone could be used for retrieval. For example, it is impossible to imagine requests such as "cotton mill workers, turn of the century," or "civil rights demonstrations," being answered using query pictures. Used in conjunction with verbal indexing systems, however, query pictures have the potential to improve image access in certain cases.

A Study of User Queries

Previous Studies. Some researchers have attempted to gain information about users' visual information needs by collecting and analyzing user queries. Two of the most significant studies are those of Keister and of Enser and McGregor.³⁰

Lucinda Keister reports that the automated retrieval system for the National Library of Medicine's Prints and Photographs Collection is "based on a detailed analysis of user queries."³¹ She describes a study involving the analysis of one year's worth of user queries from a reference log. Because the queries were reconstructed from "staffers' cryptic notes," no formal analysis of the query terms themselves was possible, but the queries were analyzed for

²⁸ Cawkell, "Picture-queries and Picture Databases," 411.

²⁹ A.E. Cawkell, *A Guide to Image Processing and Picture Management* (Aldershot, Hampshire, England: Gower, 1994), 137–41.

³⁰ Keister, "User Types and Queries," 7–19; Peter G.B. Enser and Colin G. McGregor, *Analysis of Visual Information Retrieval Queries. British Library R&D Report 6104* (London: British Library, 1993).

³¹ Keister, "User Types and Queries," 9.

the “concepts that lay behind” the user requests. Users’ professions were connected to the queries from memory.³²

One conclusion reached by this study is that one-third to one-half of all requests are “image construct queries,” or verbal descriptions of an image either remembered or imagined.³³ In Keister’s words, “Although isolated terms in these queries may be topical, the concept behind them is a visual construct, e.g., ‘people racing in wheelchairs,’ or ‘surgeons standing.’” She argues that such requests cannot be met without greater indexing at the pre-iconographical level. The study also found that picture researchers tend to have visual requirements for their images, and to use “graphics jargon” in their requests.³⁴

Enser describes a study of 2,722 user requests for picture material recorded by the Hulton Deutsch Collection in England. Details related to subject, customer, and the number of pictures retrieved were recorded by the Hulton on a standard form. These requests were then analyzed according to whether the subject was *unique* (e.g., “the paddlesteamer *Medway Queen*”) or *nonunique* (e.g., “paddlesteamers”), and whether or not it was *refined* in terms of “time, location, action, event, or technical specification.”³⁵

In terms of these categories, it was found that the majority of requests (69 percent) were for unique subjects, and that 52 percent were refined in some way. It was noted that one refiner, “time,” was used in 34 percent of all queries. In examining the relationship between query categories and user types, it was seen that requests for unique subjects were especially high among newspaper and magazine publishers, and that refiners were used most often by magazine publishers and advertising agencies.³⁶

In a second part of Enser’s study, an attempt was made to match a sample of the queries with one of the classification systems (Gibbs-Smith) used by the Hulton Deutsch. It was found that “the Gibbs-Smith scheme can function only as a blunt pointer to regions of the Hulton collections where pertinent material might be co-located.” This was partly due to the frequent use of refiners in requests, as discussed above.³⁷ Enser concludes pessimistically that “[the] findings of the investigation reported here lend support to the view that the subject indexing of commercial photographic material is of low utility,” and that for the Hulton Deutsch, “extremely heavy dependency on the

³² Keister, “User Types and Queries,” 10.

³³ Keister, “User Types and Queries,” 13.

³⁴ Keister, “User Types and Queries,” 9.

³⁵ Peter G.B. Enser, “Query Analysis in a Visual Information Retrieval Context,” *Journal of Document & Text Management* 1(1993): 29.

³⁶ Enser, “Query Analysis,” 31–32.

³⁷ Enser, “Query Analysis,” 35.

expertise and search time availability of the picture researcher in his/her role as intermediary must continue to be the norm.”³⁸

Several researchers have stressed the importance of additional research into users’ visual information needs. Cawkell, in his discussion of the possibilities for picture databases, states: “Not a great deal is known about the kind of questions that users are likely to put to a picture collection. What kind of users and what kind of collection are we talking about? . . . [T]here is virtually no analysis of the kind of questions asked or the success of the system in responding to those questions in any of the published papers.”³⁹ In the introduction to their own study of user queries, Enser and McGregor state that, “A comprehensive inspection of the literature of image database and retrieval applications leaves the authors unaware of any systematic study of user demand for visual information sources.”⁴⁰

Shatford Layne points out that while many authors argue for the importance of indexing various attributes, their conclusions are justified by “a combination of theory, logic, and anecdotal evidence.” In her words:

[I]t would be helpful if there were studies of image materials that provided quantitative evidence of the relative usefulness of the various attributes of images in providing access to those images. Unfortunately, use or user studies of image materials are not numerous, and of those studies that exist, very few contain information as to which attributes of image materials it would be useful to have indexed.⁴¹

Although Keister and Enser each studied user queries, they did not analyze them in a way that would shed light on the above issue. In grouping most subject terms and attributes into a single class called “refiners,” Enser passed over a valuable opportunity to compare the usefulness of different elements for subject access. For example, Enser states that, “The use of refiners. . . by advertising agencies and magazine publishers is especially noteworthy, this often reflecting the need for iconic material capable of inducing a desired response, e.g. *crying, distress, must be over 16, good focus on individual*.”⁴² It would be interesting to know how often such emotional or visual qualities were requested, as well as the relative frequency of other types of attributes.

The Current Study. The current study analyzed user queries at two institutions maintaining collections of historical photographs. It attempted to provide answers to the following questions:

³⁸ Enser, “Query Analysis,” 38.

³⁹ Cawkell, “Picture-queries and Picture Databases,” 418.

⁴⁰ Enser and McGregor, *Analysis of Visual Information Retrieval Queries*, 1.

⁴¹ Shatford Layne, “Some Issues in the Indexing of Images,” 587.

⁴² Enser, “Query Analysis,” 32.

- 1) What percentage of queries involves terms describing generic subject matter (roughly corresponding to the pre-iconographical level of description)?
- 2) Are there attributes of images that are not being indexed but would be useful for retrieving images? For example, what percentage of requests indicates emotional or expressional requirements for images (e.g., humor, drama)? Visual requirements (e.g., interior/exterior, color)? Other requirements?
- 3) Which categories of terms (e.g., generic subject terms, specific subject terms, time, place, genre, etc.) are most commonly used in queries?

Setting for the Current Study. Patron queries were collected over a four-month period at two institutions: the Photographic Archives of the North Carolina Collection (Wilson Library, University of North Carolina at Chapel Hill), and the photographic section of the North Carolina State Archives in Raleigh. Both collections are historical and focused on North Carolina.

Patrons of these institutions represent book, magazine, newspaper, and journal publishers; local archives, museums, and historical societies; government agencies; academic departments at colleges and universities; advertising agencies; video or television production companies; and organizations such as restaurants, churches, Chambers of Commerce, volunteer organizations, and others. In addition, many of the patrons (approximately 30 percent at each institution) are individuals seeking images for their personal research. Approximately one-third of the patrons of the North Carolina Collection are affiliated in some way with the University of North Carolina, whereas a greater percentage of patrons of the State Archives represent local archives, museums, historical societies, and government agencies. (See Table 1 for a profile of the patrons participating in this study.)

In both institutions, patrons are seeking images primarily for publication or exhibition purposes. Other reasons mentioned included the need to see architectural details of a demolished building, the desire to see a picture of a relative, and research for a theatrical production.

Data Collection. Patrons were asked to write their request(s) for image material on a form designed to elicit all aspects of their visual information needs. Two different forms were used during the course of this study, one containing a revised version of one of the two questions (see Appendices A and B).

The first question on the original form asked the patron to describe in his/her own words as clearly and completely as possible the image(s) he/she was looking for. On the revised form, this question was expanded to specifically prompt the user for information pertaining to the subject, place, time, visual characteristics, emotional or subjective qualities, or other requirements of the image(s) sought. The objective in both cases was to gather as much

information as possible about the user's visual information needs (as perceived by the user when initially making the request).

The revised question was created in an attempt to mirror more closely the reference interview and thereby elicit more detailed information about the user's image needs. It appeared, however, that the expanded form made little if any difference in the wording of the requests. A patron desiring a picture of a specific building, person, or subject (e.g., "Ray Anderson, sky-diving, 1930s") generally wished to see *any* images matching these subject requirements, and more detailed description of the desired subject matter was not possible. A visual examination of the images was necessary in order for any other needs or preferences to be expressed. Because the revised form had no noticeable effect on the nature of the queries, the data from both forms were combined for this study.

The second question on each form asked the user to describe the capacity in which he/she was visiting the archives, e.g., as a picture researcher for an alumni magazine, as a restaurant owner looking for historical photographs to exhibit, etc. This information was used to develop the profiles of user types for each institution.

Queries received by telephone were recorded on the same form by the archivist, with the archivist attempting to gather as much information as possible about the image needs of the patron. Queries received by letter or fax were analyzed exactly as they were received.

Data Analysis. Each query was carefully examined to determine which categories of attributes and subject terms represented the subject matter of

Table I. Patron Profiles

North Carolina State Archives	Number	Percent
Personal research	21	28
Archives, museums, historical societies	20	26
Book & magazine publishers	9	12
Other organizations	8	11
Academic departments	6	8
Government agencies and offices	5	6
Advertising agencies	4	5
Video or theater production companies	3	4
Total:	76	100
North Carolina Collection	Number	Percent
Book, magazine, newspaper, and journal publishers	17	33
Personal research	16	31
Other organizations	9	18
Academic departments	4	8
Museums & historical societies	3	6
Advertising agencies	1	2
Television production companies	1	2
Total:	51	100
Patrons affiliated with UNC	16	31

the query as it was expressed by the patron. A list was made of the categories used in the queries as they naturally evolved. Ultimately, the list included all categories of terms needed to describe the subject matter of the queries received. Aspects which did not fit into this list were also noted.

Some interpretation of patrons' words was required in order to determine which categories most appropriately represented the visual need expressed by the query. For example, a request for pictures of "nineteenth-century black university employees" was counted as a "Corporate Name" query, even though the University of North Carolina was not named explicitly. This query was also counted under the categories of "Persons" (African-Americans, employees) and "Decade."

For each category on the list, a tally was made of the number of queries employing that type of term. Queries using terms referring to time periods longer than a decade (e.g., "1850-1900," "1800s"), were counted in the "Decade" category, on the assumption that the patron could search multiple decades as needed. Similarly, queries describing regions of North Carolina encompassing multiple counties were included in the "County" category.

Queries from the two institutions were analyzed separately, in order to determine if the different patron profiles produce significantly different types of queries. The results for the two institutions were also combined.

Results. Fifty-one forms were gathered at the North Carolina Collection, representing 100 queries. Seventy-five forms were collected at the State Archives, representing eighty-seven queries. Some forms contained one query, whereas others listed two or more separate queries. (See Table 2 for a summary of the numbers and percentages of queries employing various categories of terms.)

Subject terms were used far more often than any other class of attributes in requests for images. Eighty-six percent of all queries included subject terms, with 57 percent including generic terms and 42 percent including specific terms. Some examples of generic and specific subject terms contained in requests are given below:

Generic terms:

- fire brigades, families, students (Persons)
- waterfront, downtown (Geographical features)
- telegraphs, plank roads, snowman (Objects/things)
- weaving, football, hunting, rallies (Activities)
- segregation, racism (Concepts)

Specific terms:

- Jesse Helms, Ava Gardner (Personal Name)
- J. C. Brantley Drug Store, UNC Press (Organization Name)
- New Hope Valley, Great Smoky Mountains (Geographical Name)

Table 2. Summary of Results

	NCC (100 queries)		State Arch. (87 queries)		Total (187)	
	#	%	#	%	#	%
Persons	27	27	9	10	36	19
Geographical	1	1	1	1	2	1
Objects/things	34	34	23	26	57	30
Activities	14	14	16	18	30	16
Concepts	4	4	3	3	7	4
Total generic:	61	61	46	53	107	57
Personal Name	17	17	15	17	32	17
Organization Name	20	20	13	15	33	18
Geographical Name	1	1	4	5	5	3
Object Name	2	2	1	1	3	2
Building Name	5	5	9	10	14	7
Event Name	2	2	8	9	10	5
Total specific:	38	38	41	47	79	42
Total subject:	87	87	74	85	161	86
Year	7	7	11	13	18	10
Decade	33	33	32	37	65	35
Total time:	37	37	43	49	80	43
Street	1	1	1	1	2	1
City	11	11	25	29	36	19
County	6	6	12	14	18	10
Total place:	16	16	34	39	50	27
Landscape	0	0	2	2	2	1
Street scene	2	2	1	1	3	2
Aerial view	0	0	2	2	2	1
Portrait	5	5	6	7	11	6
Group portrait	1	1	0	0	1	1
Total genre:	8	8	11	13	19	10
Interior/exterior	2	2	2	2	4	2
B&W/color	4	4	4	5	8	4
Detail	2	2	0	0	2	1
Total visual:	8	8	6	7	14	7
Format	1	1	0	0	1	1
Reproduction	10	10	1	1	11	6
Total physical:	11	11	1	1	12	6
Creator/provenance	4	4	3	3	7	4

Great Dismal Swamp Canal, Old Well (Object Name)

Memorial Hall, Tryon Palace (Building Name)

The Lost Colony, Civil War (Event Name)

In contrast, only 4 percent of all queries from either institution specified the creator of the image. At the North Carolina Collection, there were two requests for reproductions of John White's drawings, one request for photographs by Bayard Wootten, and one request for photographs taken by the UNC Press. At the State Archives, three requests specified photographs pub-

lished in a particular newspaper. No other requests specified creator or provenance in any way.

Specification of a particular physical form for an image was similarly infrequent. Only 6 percent of all requests contained any reference to the physical form of the image, with one patron requesting a daguerrotype image and ten others requesting reproductions from other media (such as drawings, maps, etc.).

Likewise, genre and visual terms were seldom used in requests. Genre terms were specified in 10 percent of the queries, with the categories of landscape, street scene, aerial view, portrait, and group portrait being mentioned. The visual attributes of interior versus exterior, black and white versus color, and detail were mentioned in 7 percent of the queries.

After subject terms, time was the second most frequently mentioned attribute, with 43 percent of the queries specifying date in some way. Place was next in importance, with 27 percent of all queries naming a street, city, county, or region. (Because both collections focus on North Carolina, "state" was not a relevant category for analysis.)

Eight queries also contained subjective requirements that did not fit into any of the above categories. These are listed below:

- "small town feel"
- "nice Christmas gift"
- "looks old and historical. . .recognizable" [for before and after comparisons]
- "images old people can relate to"
- "humor, odd compositions"
- "old photo for gift"
- "vibrancy or vigorous character in the image, suggestive of agency and self-confident political ambition"
- "stark realism—photographs that would give clues about state of mind"

There was little difference in the rankings of the attribute classes between institutions. Although generic subject terms were used relatively more frequently at the North Carolina Collection, this class of terms ranked first in both places. Street, city, and county were used relatively more often in requests at the State Archives, however, place terms ranked fourth for both institutions. Even within classes of categories, terms that were rare at one archives (such as "Geographical features," "Concepts," and "Street") were also rare at the other.

Conclusion

The use of generic terms does not necessarily constitute description at the pre-iconographical level, as these words generally require more than "ev-

everyday knowledge” for their comprehension. For example, some cultural knowledge is required to understand such generic terms as “privies” or “sharecroppers.” (A pre-iconographical description of an image depicting the latter, for example, might read simply “men working in a field.”) The distinction between pre-iconographical and iconographical levels of analysis, however, is perhaps less useful for describing ordinary images than it is for art images. What is really needed is access to images in terms of their generic content, so patrons from any discipline can find images for purposes not predictable by the cataloger. The high percentage of queries containing generic subject words indicates the importance of these terms for image retrieval.

Similarly, proper names and terms describing time and place were seen to be very important access points. Specific subject terms were used almost as often as generic terms, with personal and organization names being the categories used most often. Decade was specified more often than year, and city was specified more often than street or county, indicating the relative usefulness of these types of terms. It is important to note that place and time can be useful for retrieving images even when they are not specifically mentioned in a request.

These results concur with those of Helen Tibbo, who studied historians’ search requests in an effort to determine which categories of information should be included in historical abstracts.⁴³ Subject words were by far the most commonly used type of term, while words indicating place, time, and proper names were found to be next in importance. It is not surprising that the results of the current study would be similar, given that the image description is, in this case, an abstract of a historical document (the historical photograph). For other types of image collections (e.g., art, scientific), different categories of terms might have greater importance.

Although patrons are often interested in the expressional qualities of images, the results of this study give little indication that a greater amount of indexing of these qualities would significantly increase access. Although emotional or subjective qualities were specified by eight patrons, it is difficult to imagine in any of these cases how the desired qualities could have been indexed. In order to determine which images meet such criteria, visual examination of the images or image surrogates is required.

Likewise, even though images were primarily being sought for the purposes of publication and exhibition, terms indicating visual requirements (e.g., landscape versus portrait orientation, or color versus black and white) were rarely used in the queries. Only 7 percent of the queries indicated any kind of visual requirements, and these were limited to interior versus exterior,

⁴³ Helen R. Tibbo, *Abstracting, Information Retrieval and the Humanities: Providing Access to Historical Literature* (Chicago: American Library Association, 1993).

black and white versus color, and detail requirements. This does not mean that such requirements or preferences do not exist, but only that they are not always stated or recognized at the early stages of the search process. On the other hand, requests for pictures from the archives in this study were generally extremely subject-driven, with the patron wishing to see any material that could be produced on the subject of interest. Queries containing visual terms may be more common at institutions whose patrons include a greater number of graphic designers.

Although genre and physical attributes were mentioned in only 16 percent of the queries, these terms are absolutely necessary for certain types of retrieval (e.g., the request for a daguerrotype image). Similarly, it would be difficult to fill requests for landscapes, street scenes, or aerial views without indexing genre terms. Again, whether such indexing is worthwhile for a particular institution will depend on that institution's patrons and their visual information needs.

Perhaps surprisingly, image creator was mentioned in only 4 percent (7 out of 187) of the queries. This is particularly significant in light of the fact that most archives provide access to their materials primarily through provenance, thereby necessitating some knowledge of an image's origin or use if no additional subject access is provided. In the archives studied, patrons were generally ignorant of the collections held and sought images individually. In fact, not one of the requests recorded during this study involved an interest in studying images in the context of a collection. While maintaining the context in which images were created or used is necessary to preserve their evidential value, it is clear that few patrons are presently using images as primary source documents. A study of how patrons use images, and the implications for archives, would be interesting and useful.

Ultimately, the needs of each particular institution must be independently addressed in determining the depth of indexing required for a collection. What works for one institution may not work for another, given that the nature of its holdings and therefore the needs of its patrons are possibly very different. An important point made by members of the UC Berkeley Digital Image Access Project is that, as desirable as item-level cataloging may be, the greatest obstacle facing access to many image collections is simply the sheer volume of images waiting to be processed. For example, Berkeley's Bancroft Library has collections containing approximately 3.5 million images, "virtually none of which could be considered fully cataloged by existing standards for bibliographic control." Approximately \$400 million and 400 years of cataloging staff time would be required to catalog all of these images at the item level.⁴⁴

⁴⁴ "SGML: California Heritage Digital Image Access Project," <<http://www.sil.org/www/sgml/berkenh.html>> (accessed 27 Feb. 97).

Perhaps the most efficient way to provide access is to combine indexing of the most commonly requested attributes with the ability to visually examine many images. As Shatford Layne states, "Rather than devoting time to extraordinarily detailed or complicated indexing, or to elaborate parsing schemes that refine verbal searches, it might be better to concentrate on indexing the basic elements of an image and rely on scanning. . .to make the fine distinctions."⁴⁵ Digital image databases that combine the presentation of thumbnail representations of images with verbal indexing schemes provide the possibility for such access.

It is interesting to note that although there is always the problem of indexing consistency, the categories of terms used most often by patrons in this study are also among those easiest to agree upon. The majority of patron requests indicated the image subject in terms of generic classes of persons, things, and activities depicted—items generally recognizable with only everyday knowledge and more likely to be agreed upon than a picture's expressional or conceptual qualities. The other most common type of request described the image subject using personal name, organization name, building name, or other proper nouns. These terms refer to unique items and, if identifiable, are not subject to subjective interpretation.

Regardless of the time and expense involved, without subject access, many requests for images could not be answered. The fact that the overwhelming majority of the patrons in this study sought images by subject indicates the importance of these terms for retrieval. Any amount of subject indexing, even of only the main subjects of photographs, could only improve access.

⁴⁵ Shatford Layne, "Some Issues in the Indexing of Images," 586.

Appendix A. First Form Used for Query Recording

My name is Karen Collins. I am a graduate student in the UNC-CH School of Information and Library Science. I am collecting information about user requests in an effort to learn ways to improve access to photographic collections. This study is part of my master's paper research. **Please note that by answering the questions below you are consenting to participate in this study. Your participation is voluntary.** Note that your name will not be connected with this form in any way. Your assistance is greatly appreciated.

Karen Collins (Principal Investigator) 929-2293, Helen Tibbo (Faculty Advisor) 962-8063, Frances Campbell (Institutional Review Board Chair) 966-5625

1. Please describe in your own words, as clearly and completely as possible, what photographs you are looking for.

2. Whom are you representing?

- ☐ UNC department specify: _____
- ☐ self (personal research)
- ☐ publisher specify: book/journal/magazine/newspaper
- ☐ advertising/design company
- ☐ video or film producer
- ☐ other specify: _____

Appendix B. Second Form Used for Query Recording

My name is Karen Collins. I am a graduate student in the UNC-CH School of Information and Library Science. I am collecting information about user requests in an effort to learn ways to improve access to photographic collections. This study is part of my master's paper research. **Please note that by answering the questions below you are consenting to participate in this study. Your participation is voluntary.** Note that your name will not be connected with this form in any way. Your assistance is greatly appreciated.

Karen Collins (Principal Investigator) 929-2293, Helen Tibbo (Faculty Advisor) 962-8063, Frances Campbell (Institutional Review Board Chair) 966-5625

1. Please describe in as much detail as possible what photographs you are looking for. Leave blank any categories that are not applicable.

subject?

place?

time?

visual characteristics (e.g., landscape vs. portrait orientation, black & white vs. color, etc.)?

emotional/subjective qualities?

other requirements?

2. Whom are you representing?

☐ UNC department specify: _____
☐ self (personal research)
☐ publisher specify: book/journal/magazine/newspaper
☐ advertising/design company
☐ video or film producer
☐ other specify: _____