Hannah M. Noll. Where Google Stands on Art: An Evaluation of Content Coverage in Online Databases. A Master's Paper for the M.S. in L.S degree. April, 2008. 43 pages. Advisor: Jeffrey Pomerantz

This study evaluates the content coverage of Google Scholar and three commercial databases (Arts & Humanities Citation Index, Bibliography of the History of Art and Art Full Text/Art Index Retrospective) on the subject of art history. Each database is tested using a bibliography method and evaluated based on Péter Jacsó's scope criteria for online databases. Of the 472 articles tested, Google Scholar indexed the smallest number of citations (35%), outshone by the Arts & Humanities Citation Index which covered 73% of the test set. This content evaluation also examines specific aspects of coverage to find that in comparison to the other databases, Google Scholar provides consistent coverage over the time range tested (1975-2008) and considerable access to article abstracts (56%). Google Scholar failed, however, to fully index the most frequently cited art periodical in the test set, the *Artforum International*. Finally, Google Scholar's total citation count is inflated by a significant percentage (23%) of articles which include duplicate, triplicate or multiple versions of the same record.

Headings:

Bibliographic Databases – Evaluation

Electronic Information Resources

Google Scholar (Web site) – Evaluation

# WHERE GOOGLE SCHOLAR STANDS ON ART: AN EVALUATION OF CONTENT COVERAGE IN ONLINE DATABASES

by
Hannah M. Noll

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2008

Approved by

_____

Jeffrey Pomerantz

Table of Contents

# Introduction

Google Scholar's slogan, "Stand on the Shoulders of Giants," derived from a quotation by Isaac Newton, is based on the philosophy of the service that "much of scholarly research involves building on what others have already discovered" (See "Google Scholar Help", at http://scholar.google.com/intl/en/scholar/help.html). With a single search box serving as the portal to millions of scholarly resources, Google Scholar (GS) presents itself as a formidable competitor to long-standing citation indexes and commercial databases, such as the ISI Citation Index. Google Scholar has generated praise for its potential to 'democratize' the public's access to academic research (Noruzi, 2005). Yet critics also claim that the citation search engine suffers from severe gaps in subject coverage and lacks the advanced search features to retrieve relevant information effectively (Jacsó, 2005a, 2005b).

The scope of the citations indexed by Google Scholar can be difficult to pin down because unlike a traditional citation database, it is crawler-based like a search engine, and provides results extracted from the millions of web sources that it indexes. Google Scholar has gained access to scholarly web content through its development of partnerships with the creators of these web sources, which range from publishers to academic institutions, professional societies to preprint repositories (See "About Google Scholar", at http://scholar.google.com/intl/en/scholar/about.html). Much to the disappointment of librarians and critics, however, Google Scholar has not divulged the

names of these partners and has released only vague details about the scope of its database.

The secrecy surrounding Google Scholar's holdings has created a genuine need for content evaluation of the subject areas purportedly included in the scholarly search engine. What exactly does Google Scholar mean when it describes its subject coverage as 'scholarly literature from all broad areas of research' (See 'Google Scholar Help' page). One way to get a sense of the subject scope of Google Scholar is to perform test searches in the database. Numerous content evaluation studies have adopted this approach to determine Google Scholar's strengths in specific subject areas. Initially reviewers focused on Google Scholar's coverage of medical and scientific subject areas (Abram, 2005; Shultz, 2007), but in recent studies critics have begun to explore the content available in the social sciences and multi-disciplinary subject areas (Gardner & Eng, 2005; Callicott & Vaughan, 2005; Walters, 2007).

Fine arts are hardly at the forefront of database content studies. Yet the potential, and inevitable, use of Google Scholar among undergraduate students, and art scholars alike, requires consideration. This study looks at how Google Scholar measures up to three major art databases (Art Full Text/Art Index Retrospective, Arts & Humanities Citation Index (A&HCI), and the Bibliography of the History of Art (BHA)) on the subject of art history. To get an understanding of where Google Scholar stands on art, this study will test and compare each database for article citations authored by a pre-selected set of art historians. The methodology for this study applies aspects of Jacsó's database evaluation criteria with Tenopir's definition of the bibliography method, to the subject coverage of art history in online databases. The study is also influenced by

William H. Walters' evaluation of commercial databases and Google Scholar through sample tests for a 'core set of articles' on the multi-disciplinary topic of later-life migration (2007). A 'core set' of art historians is culled from the *Dictionary of Art Historians* (http://dictionaryofarthistorians.org/), which compiles the biographies of art historians mentioned in major art historiographies. The 472 articles published by this 'core set' of art historians represents a body of significant or relevant scholarly literature in the discipline of art history.

A perfect comparison of Google Scholar and the three commercial databases is impossible, because the search mechanics differ significantly between the databases (Shultz, 2007). But as critics point out, Google Scholar *will* be popular, and librarians must be prepared to know its strengths and weaknesses (Gardner & Eng, 2005). This study will compare the overall coverage across databases, as well as provide an analysis of general scope characteristics found in the databases, to reveal a sense of how shaky or how solid Google stands on art history literature. In doing so, this study will add to the growing literature that has helped to further demystify the content coverage of Google Scholar.

## Literature Review

### *Jacsó & Database Scope*

Péter Jacsó, a professor of Library & Information Science and database critic, has written extensively on databases and Google Scholar in his columns for *Online Information Review*, *Online* and *Galenet*, in addition to numerous journal publications and conference talks (see http://www2.hawaii.edu/~jacso/ ). Virtually every article written about Google Scholar cites Jacsó's research. His contributions to the literature

cannot be understated. In addition to his rank as a distinguished Google Scholar critic, his book on textual CD-ROM and web databases serves as a comprehensive textbook for library professionals interested in evaluating the content of any database.

Jacsó describes a database's scope as coverage related to specific criteria like size, composition, extent of coverage of source documents or objects, currency, coverage by language or geographic region, and subject area (2001, p.15). Some of these criteria are highly interrelated. He explains that the size, extent of source coverage, and subject area of a database are especially intertwined. This study will adhere to Jacsó's definition: "In the broadest sense, the scope of the database defines the coverage" (2001, p.15). The scope of a database can be ascertained by the combination of a database provider's description and a reviewer's test searches. He warns that a provider's description should be taken with a grain of salt and that test studies are needed to gain a more complete understanding of a database's scope (Jacsó, 2001).

### *Google Scholar & Database Scope*

The Google Scholar site gives this skeletal description about its scope:

> Google Scholar covers peer-reviewed papers, theses, books, abstracts, and other scholarly literature from all broad areas of research.
> (See "Google Scholar Help", at
> http://scholar.google.com/intl/en/scholar/help.html).

Unlike many other commercial databases, Google Scholar has not released much information about what is contained in its database. Many reviewers find this to be a point of major concern (Neuhaus, 2006; Schroeder, 2007; Jacsó, 2005a, 2005b; Burright, 2006) and one reviewer goes so far as to demand that Google Scholar developers stop pleading the "Fifth" and release this source information to the public (Jacsó, 2008).

While Google Scholar appears to include a wealth of some source content, like OCLC's Worldcat holdings (Callicott & Vaughn, 2005; Tenopir, 2005), it has only recently begun to index some of the most important publishers in the medical and scientific disciplines, Elsevier and American Chemical Society (Jacsó, 2008).

### *Size*

Many commercial databases, like those used in this study, give an approximate record count in the description of the database. For example, the Arts & Humanities Citation Index (A&HCI) claims in its description to include over 2.5 million records in its database (See "Web of Science Help Page", at http://images.isiknowledge.com.libproxy.lib.unc.edu/help/WOS/h_database.html). Google Scholar, on the other hand, has not released any estimation of the size of its database, though Scholar engineer and creator Anurag Acharya has said in an interview that Google Scholar is "pretty large" (Quint, 2007). It is almost impossible to explore the size of Google Scholar with test searches because the software prevents any reliable way to retrieve that information (Jacsó, 2005a). Jacsó reports that despite hundreds of hours spent testing Google Scholar; he is not closer to figuring out its size, because the usual bibliometric techniques used to test the size of a commercial database are confounded by GS's improbable hit counts and inconsistencies. As an example he explains that a search by 1455-2005 as year range retrieves 1,388,000 hits, whereas a search for the word "protein" yields 7,390,000 results (Jacsó, 2005c).

### *Composition*

The size of a database, or the total record count, must be considered within the context of a database's composition (Jacsó, 2001). Composition refers to the makeup of

the database not just based on numbers, but based on the characteristics of each individual record. This includes scope characteristics like article or document type, language of publication or subject area of the article. A history database might include 10 million records, but if its overall composition does not include any primary source documents, it might not be useful or relevant for certain types of research. In raw numbers, Google Scholar retrieves more records than other commercial databases (White, 2006). On an item by item comparison, however, the retrieved records in GS are inflated with duplicate records and citations to web documents that might not be considered 'scholarly' (e.g. digital pamphlets, pre-prints, etc.) (Jacsó, 2006; Walters, 2007). In addition to duplicate records, Google Scholar includes multiple versions of the same article and this may lead to what Jacsó calls a 'false impression' of the size and composition of the database (Jacsó, 2005b).

### *Date Coverage*

The date ranges included in Google Scholar are also unknown. Jacsó jokes that coverage may begin as early as 868 AD "in case Google, Inc. got the right for the Google Print project to digitize the Buddhist Diamond Sutra, believed to be the first printed book," (2005a, p. 1541), but no official date range is included in the Google Scholar information pages. In terms of currency, Google Scholar does not include any details about how often the database is updated. One study found that the upload frequency for the database could be delayed as long as 12-15 weeks for new records added to the database (Neuhaus et al., 2006) and another found a two-month lag time in the coverage of the publication *Nature* (Burright, 2006). Currency is not necessarily a priority for arts

and humanities research, but other researchers, like those in medical and scientific disciplines, may require access to the most up-to-date information on a subject.

### *Language & Geographic Coverage*

Google Scholar does not describe its language or geographic coverage, but test studies have targeted this aspect of GS's scope and have returned mixed conclusions. Neuhaus et al. observed a significant English language bias (2006), while another study found that Google Scholar was stronger in European language content than other commercial databases, like Web of Science (Noruzi, 2005). Mayr & Walter discovered that Google Scholar retrieved results from many foreign language journals, but the majority of the results did not provide full text links to the foreign language abstract or full-text article (2007). Currently Google Scholar does not index languages with 16-bit character scripts, such as Arabic, Chinese or Japanese (Noruzi, 2005). Commercial databases in the USA typically have an English-language bias, but Google Scholar has shown a significant strength in its language coverage and has been sending its web crawlers into increasingly international digital document collections (Jacsó, 2008).

### *Databases & Content Evaluation*

"The systematic evaluation of the content of databases is one of the core activities of database quality assessment," (Jacsó, 1997, p.233). Jacsó points out that content, defined as the information created by the author, is at the core of database assessment, because without relevant information on a subject, the functions of the database do not matter. As librarians and information professionals often act as the intermediaries between databases and their users, Jacsó notes, it is important that the content coverage of

a database be thoroughly evaluated for its particular strengths and weaknesses in a subject area (1997).

Carol Tenopir, Professor of Information Science and author of the monthly "Online Databases" column in *Library Journal*, distills the two most common methodologies used in evaluating database coverage on a specific subject (1982). In the first method, the bibliography method, a reviewer uses a bibliography in a specific subject area (e.g. Renaissance Art) to test and compare the bibliographic coverage across databases. The database with the largest percentage of citations from the bibliography is then considered to have better content coverage than a database which includes a smaller percentage of the citations.

In the second method, the 'subject profile' method, a comprehensive list of terms is compiled to represent a broad overview of a subject. These terms are then tested in each database, and whichever database retrieves the most citations would be considered to have more extensive coverage on a particular subject. Tenopir applies both the bibliography method and the subject profile method to the subject of volcanology across two geology databases, the Geological Reference File and GeoArchive. She finds that both methods produced roughly the same conclusions. While Tenopir concludes that the bibliography method proved more time-consuming and costly for the database reviewer, she acknowledges that the bibliography method was the only way to determine how much of the subject literature is included in the database (1982).

Jacsó and Tenopir have contributed a significant body of work to the literature on database content evaluation. They both acknowledge that current methodologies are indebted to earlier comparison studies of print indexing and abstracting periodicals. In

one such influential study, Martyn and Slater tested major abstract journals for coverage of scientific and technical information. The researchers used what Tenopir would later define as a 'bibliography method' to compare coverage across individual abstracts journals (Martyn & Slater, 1964). F.W. Lancaster also used a bibliography method in his early evaluation of print indexes in the medical sciences. He asserts that a comprehensive bibliography on a subject can be used effectively to test an index for coverage in a specific area. To avoid bias, however, Lancaster warns readers that the bibliography should not be compiled from the index that is being tested (1971).

### *Google Scholar & Content Evaluation*

Many studies about Google Scholar have focused on the service's effectiveness for search and retrieval; far fewer studies have analyzed the subject content or topical coverage that defines the database's scope (Walters, 2007).

> Getting the big picture of the composition, size and dimension of databases is important for an overall sense of the databases. Getting little pictures for well-defined, unambiguous searches by subjects, authors, journals…brings the issue to a human scale. Together, the little pictures may corroborate the validity of the big picture—if the numbers are taken with a grain of salt.
> *Péter Jacsó, on his comparison of Google Scholar, Web of Science and Scopus* (2005a, p.1542).

The big picture of Google Scholar's subject coverage comes from the numerous content evaluation studies that have emerged since GS's 2004 release. In these studies, researchers have applied the traditional criteria for database scope to commercial databases, and then applied the same criteria to Google Scholar for comparison. Side-by-side comparisons in database studies can help put the content evaluation in perspective (Jacsó, 2001). These studies are heavily influenced by the methodologies of the subject profile and bibliography methods that Carol Tenopir defined in her 1982 article.

So, what subject disciplines does Google Scholar cover? This question was posed to Google Scholar engineer Anurag Acharya. He responded, "We believe Google Scholar covers all major disciplines" (Hughes & Acharya, 2006). His reply echoes Google Scholar's description of its subject scope on its website, which claims coverage of "all broad areas of research" (see 'Google Scholar Help'). Compare this to the description of subject coverage in another multi-disciplinary database, the Arts & Humanities Citation Index (Table 1).

Table 1
Disciplines Included in A&HCI

| Disciplines | |
|---|---|
| Archaeology | Linguistics |
| Architecture | Literary Reviews |
| Art | Literature |
| Asian Studies | Music |
| Classics | Philosophy |
| Dance | Poetry |
| Folklore | Radio, Television & Film |
| History | Religion |
| Language | Theater |

*Source:* http://images.isiknowledge.com/help/WOS/h_database.html

This lack of subject coverage description in Google Scholar has inspired reviewers to design test studies that further examine the subject coverage of Google Scholar.

In a study that compares the subject coverage of computer science across Web of Science, Google Scholar and Citeseer, one reviewer employs a modified bibliography method to compare database's coverage of publications by a single author, computer scientist and mathematician Michael Rabin. Google Scholar retrieved a considerable percentage of the test citations. Google Scholar only indexes citations from online sources, but the reviewer found a surprising strength in the back issue citations indexed from JSTOR and Project Euclid (Bar-Ilan, 2006). Jacsó uses an author-centric test for

works using Eugene Garfield as a cited author in his review of Google Scholar, Elsevier's *Scopus* and Web of Science. He finds that Google Scholar could retrieve results by Garfield, but lacked the sophisticated search options that Scopus and Web of Science have to find works with Garfield as a cited author (2005a).

Several reviewers used indexes from selected journal lists to test Google Scholar for content in scientific disciplines. Mayr and Walter took five different lists from open access journals, international scientific journals, and journals from a German social science database to get a macroscopic view of scientific coverage in GS. In terms of transparency and completeness, Mayr and Walter conclude that GS, in its current Beta form, is an unsatisfactory competitor among scientific databases (2007). Christianson used a similar method when she took a journal list for relevant articles in Ecology, selected 840 citations, and tested Google Scholar for its subject coverage in ecology. The study found that Google Scholar indexed 57 to 77% of the 840 citations (2007).

A majority of the initial reviews of Google Scholar characterized the subject scope of the database to be fairly strong in the scientific and medical literature (Wlekinski, 2005; Abram, 2005; Gardner & Eng, 2005), but severely lacking in the arts and humanities (Neuhaus et al., 2006). Neuhaus et al. found in his comparative analysis of 47 databases that GS included only 10% of the sample humanities articles, while it included 76% of the sample articles in the medical and scientific literature (2006). Other authors' reviews and studies tend to corroborate with the evidence that Google Scholar's bent is towards scientific literature (Wlekinski, 2005; Abram, 2005; Gardner and Eng, 2005), though Callicott and Vaughn found that Scholar's coverage of arts and humanities measured slightly more favorably against other subject databases (2005). In a

comparison of social sciences literature in Google Scholar, ISI, PsychINFO and Elsevier, Gardner and Eng found GS to be lacking luster in its social sciences content (2005).

A potential strength of Google Scholar is its role as a multi-disciplinary database. William H. Walters presents one of the first Google Scholar studies to incorporate the bibliography method into an analysis of relevant multi-disciplinary coverage across several databases. In his research study, Walters uses a set of 155 core articles on the multi-disciplinary topic of later-life migration. This set of articles serves as a comprehensive bibliography, much like what Lancaster describes in his article, and Walters uses it to gauge what unique and overlapping content is available in Google Scholar as it compares to subject-specific databases. The core set of articles was selected by the author based on criteria such as subject matter, importance of findings, and innovativeness of methods on the topic of later-life migration (Walters, 2007). He found that Google Scholar included 27% more core articles than the other leading database Social Science Citation Index, but that 32% of the records lacked abstracts or gave incomplete citation information (Walters, 2007).

### *Google Scholar & Search Mechanics*

While this research study focuses mostly on a subject scope comparison of Google Scholar with commercial art databases, a few functionality issues unique to GS warrant a brief overview of the literature. Google Scholar has expanded its advanced search functions over time, but some reviewers find these features to be less than satisfactory (Jacsó, 2005a; Adlington & Benda, 2006). One major point of contention is Google Scholar's lack of an authority file. Commercial databases and library catalogs have authority files for citation elements like author, publication source and date, but

Google Scholar does not provide this level of authority control. Google Scholar relies solely on the information provided by the web source that it crawls, and the citations are not regulated or reviewed by humans for verification and accuracy. The advanced search allows users to search by author, date and publication, but the absence of a real authority file limits Google Scholar's ability to function as a true bibliographic search tool (Burright, 2006).

Google Scholar does not have the sorting capabilities that are a common feature in commercial databases. These functions allow users to sort search results by variables, like author or date of publication. Sorting by date is of particular importance in scientific or medical research, which is often concerned with the currency of a publication. In addition to the absence of a reliable sort function, Google Scholar users cannot link to related or suggested subject headings (Burright, 2006). For better or worse, controlled vocabulary is also non-existent in Google Scholar.

Despite the lack of some of these search functions and added features, Google Scholar provides a speedy and massive output. Google Scholar has a built in 'stemming technology' in its search algorithm, so users do not need to apply the Boolean logic or truncation operators that other databases require (Burright, 2006). Another feature that distinguishes Google Scholar from other databases is its clever ability to spell-check a search. The enormous output of search results that GS produces is thanks in part to its ability to search, index and retrieve results based on the full text of an article, not just the title or abstract (Jacsó, 2005b). Google Scholar has adapted its search algorithm to reflect the different meaning that 'relevancy' has in a scholarly information environment. Scholar's search algorithm ranks results: "the way researchers do…weighing the full text

of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature. The most relevant results will always appear on the first page" (See "About Google Scholar"). This philosophy seems admirable, but reviewers have not found it to be entirely reliable (Walters, 2007; Burright, 2006). Further research by subject specialists is needed to determine whether the Google Scholar algorithm has achieved its claim to return relevant results in scholarly literature "the way researchers do."

### *Information-Seeking in the Arts & Humanities*

Eugene Garfield's vision of citation-enhanced searching in the sciences has had an incredible influence on the way indexes and databases are designed and used today (Jacsó, 2005a). Garfield initially created citation-enhanced searching through the Science Citation Index in 1955 and he then tacked on the Web of Science, which includes the Arts & Humanities Citation Index, twenty years later. In an article about information retrieval in arts and humanities, Garfield admits that his original interest in citation indexing sprang from its possible application to the humanities literature. In an effort to draw a distinction between the essence of research in the humanities from that of the natural sciences, he describes what he views as the information-seeking behavior of humanities scholars. According to Garfield, humanities scholars view the literature as cumulative and constantly in flux, but they are much less concerned with a work's currency than a scientist would be. Humanities scholars react and respond to published literature based on ever-changing perspectives. The type of material that is considered valuable to Humanities scholars is very different from the natural sciences. Reviews of

books, art exhibits, editorials, obituaries, essays and poems are all considered important contributions to the discipline (Garfield, 1980).

A further distinction between humanities and fine arts research is addressed by Joan Beaudoin, who carefully digests the literature on art historians' information-seeking behavior. Beaudoin points out that art historians' heavy reliance on images and objects guides their methods of research, more so than humanities scholars. While art historians have adopted new technologies in the discipline, albeit slowly, many of them use bibliographic databases to find citations, but then track down the original print sources to find the images that are absent from the online version (2005). This study recognizes that Google Scholar may be more popular among undergraduate students than among art history scholars. But it is important to understand how art historians view scholarship in the subject area of fine arts, as this may influence the types of art literature they ask students to find on their own.

### *Database content studies in the Arts*

Few studies have concentrated on database content coverage in art databases. Much of the library science literature focuses on issues surrounding visual image databases, and does not address bibliographic content across databases. One study, by Giral and Taylor, does set out to evaluate the coverage overlap in two of the largest architectural databases, Avery Index to Architectural Periodicals and Architectural Periodicals Index (1993). Giral and Taylor use a methodology based on a tradition of overlap research and inter-indexer consistency. They found the indexes held a significant overlap (71%) of the sample articles tested, but that the articles were indexed quite differently between the databases (title variations, varying access points) (Giral & Taylor,

1993). Another reviewer took articles from the art periodical *Leonardo* to investigate whether online databases mentioned or gave access to the images that accompany publisher articles in art. In this comprehensive analysis deLuise found that two of the art databases considered here (Art Full Text and the A&HCI) did not provide color plates or illustrations to accompany the text of the journal articles (deLuise, 2003).

In the Google Scholar literature, Neuhaus et al. explored subject coverage across disciplines to find that GS included only 10% of the Humanities sample set (2006). Callicott and Vaughn found that GS coverage on their sample art topic (using the search terms Andy Warhol pop art) fared slightly better with 32% of the search results deemed relevant by the researchers as suitable for use in a hypothetical research paper at an undergraduate institution (2005). As much as any other subject area, librarians must be ready to point users in the direction of the most appropriate and reliable search tools for seekers of art history information. Google Scholar's coverage of art history resources remains largely unknown. This study will set out to apply aspects of Péter Jacsó's scope criteria and Tenopir's definition of the bibliography method to the art history literature included in online databases.

## Methodology

### *Evaluation Criteria*

Multi-database comparisons serve to put some perspective on the content coverage contained across databases (Jacsó, 2001). Two art-specific databases and one multi-disciplinary citation index were selected to provide the broadest possible comparison of the subject content coverage available in Google Scholar. Subject-specific databases focus on the scholarly literature of one specific discipline. The Bibliography

of the History of Art (BHA) and Art Full Text/Art Index Retrospective are examples of two art-specific databases that index prestigious art journals, catalogs and yearbooks. Art Full Text indexes articles from 1984 to present, while Art Index Retrospective covers the literature from 1929-1984 (See Appendix). Both are listed as 'recommended databases' in the E-Research Tools page of the UNC-Chapel Hill library website. UNC describes Art Full Text as "a core database for art and art historical research," and BHA is described as "the most comprehensive bibliography available worldwide" (See http://eresources.lib.unc.edu/eid/subject.php?subjectName=Art+and+Architecture).

Multi-disciplinary databases may not include as high a concentration of subject-specific literature, but they contain a broad perspective of multi-disciplinary literature that can help searchers discover resources that may cross the boundaries of a particular discipline. In the study of art history this can be very useful. Art historians often publish in journals outside of their field, in areas like philosophy, history, literature and film. The Web of Science, which contains the Art & Humanities Citation Index, is an example of multi-disciplinary database that exposes searchers to some resources that may not be included in art-specific databases. Google Scholar does not fit into either of these database categories neatly. Like the Web of Science it allows searchers to link to and from cited works. While it is definitely not specific to just one subject, it does claim to include "the most relevant research across the world of scholarly research" (see "About Google Scholar"). As other studies have demonstrated (Neuhaus et al., 2006; Walters, 2007), evaluating Google Scholar alongside a variety of databases provides more points for comparison.

To achieve a meaningful comparison of subject coverage across databases, Jacsó reminds reviewers that it is important to remain consistent in the evaluation process. Test searches within databases should conform to types of sources and variables that are available in *all* tested databases. Language, date and publication type are examples of variables that differed between the databases selected (see Appendix). The A&HCI only indexes publications from 1975-to present. It also indexes English-language journals only. While Art Full Text and the BHA include international journal coverage, the extent of the coverage of foreign-language articles in these databases falls out of the scope of this study. Only articles written from 1975 onwards were considered, because this date range is the only one covered by all databases. While Google Scholar and the BHA include bibliographic citations to books, only articles were searched across databases (See Table 2). Jacsó talks about how certain publications, like book reviews, may inflate the purported size of a database and give reviewers a false impression about the size of a database (Jacsó, 1997). In this study, book reviews are intentionally counted as articles, because they are significant to the scholarly literature in art history (Garfield, 1980).

Table 2
Types of materials considered in this study

| **Material types included** | **Materials *not* included** |
|---|---|
| journal articles, (book, catalog or exhibit reviews), obituaries, editorials, letters, poetry, and other miscellaneous publications written by the art historian. | Monographs, dissertations, books, abstracts. |

*Bibliography Method*

In Tenopir's description of the 'bibliography' method a pre-selected subject bibliography is used to test each database for the percentage of sources covered in the

database. The bibliography represents a source of relevant and important sources on a particular subject. One drawback to this method for the database evaluator is that a bibliography must exist or be created for the purpose of the testing. Instead of using an existing bibliography, this study used an existing list of art historians known to be important and relevant contributors to art historical scholarship. These art historians appear in the *Dictionary of Art Historians* (http://dictionaryofarthistorians.org/), a biographical Dictionary of historic scholars, museum professionals, and academic historians of art:

> This database is designed to give researchers a beginning point to learning the background of major art historians of western art history. This is an index of art historians mentioned in major art historiographies.

While the *Dictionary of Art Historians* covers the biographies of over 1,000 art historians from the 16th to the 21st century, not every historian could be tested in the study. The art historians selected for this study had to conform to some of the limitations of sources included in all of the databases that were used in the study:

> *Only authors born after 1925*
> This limit was put in place to ensure that the included art historians would have publications indexed within the scope of the databases (i.e. A&CHI only indexes art periodicals from 1975 on, therefore authors born in the early 20th or 19th centuries would likely not have published works in the database).

This resulted in a list of 173 art historians. The list was further limited to:

> *Only UK or USA-born authors*
> Some of the databases covered featured foreign language holdings, but some do not. To eliminate an unfair bias for those databases that cover foreign language journals, only English language articles were included.

Additionally, the list of art historians had to be narrowed further to form a manageable test set (See Table 3).

Table 3
Art Historians tested across databases

| Name | Nationality |
|---|---|
| Gerald M. Ackerman (1928-) | USA |
| Svetlana Alpers (1936-) | USA |
| Michael Baxandall (1933-) | UK |
| Norman Bryson (1949-) | UK |
| TJ [Timothy James] Clark (1943-) | UK |
| Michael Fried (1939-) | USA |
| Rona Goffen (1944-2004) | USA |
| Linda Nochlin (1931-) | USA |
| Barbara J. Novak (1928-) | USA |
| Francis V. O'Connor (1937-) | USA |
| Griselda Pollock (1949-) | UK |
| Marvin Trachtenberg (1939-) | USA |

*Source: Dictionary of Art Historians*
*http://dictionaryofarthistorians.org/*

The author of this study is not a subject specialist in art history, and the 12 art historians selected do not reflect any sort of value judgment on their importance over other qualified art historians. Instead, the art historian set allowed this study to compile a core set of 472 articles that could be tested across the databases. Walters uses a similar methodology in his comparison of multi-disciplinary content coverage in Google Scholar and seven other databases. He compiles a bibliography of hand-picked articles deemed 'relevant' and important to studies of later-life migration (2007). The art historical relevance of each individual article could not be determined in the time frame of this study, but the large

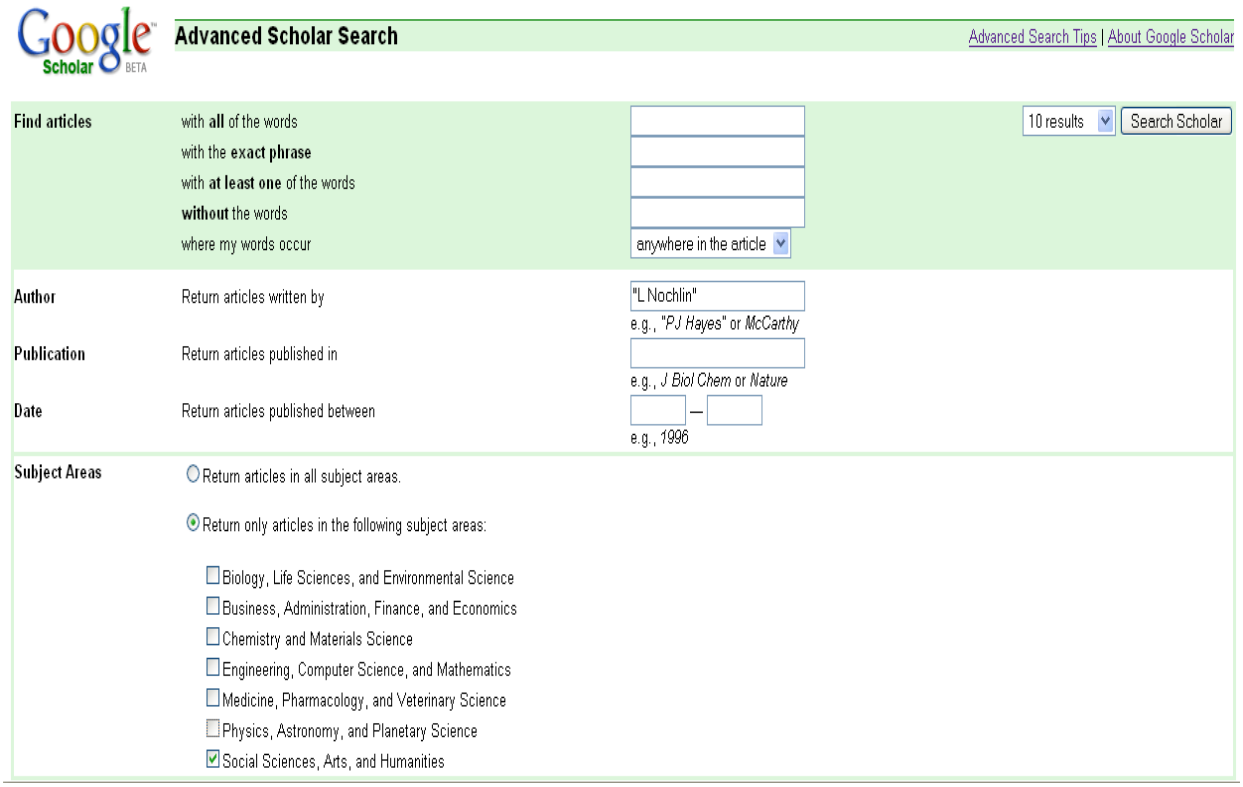
number of articles derived from the set of relevant and published art historians should provide a test set that is generally representative of literature in western art history.

The bibliography method was selected over what Tenopir defines as the second most common subject coverage methodology, the 'subject profile' method. Searching by a compiled list of subject terms did not seem appropriate for testing Google Scholar. Google Scholar lacks a system of controlled vocabulary and search results reflect in many cases a full-text search of the document, whereas traditional databases only search the title and abstract keywords of a record. Unfortunately, keyword searches in Google Scholar do not yield a manageable set of records for comparison with commercial databases. A keyword search in Google Scholar for the keywords feminism and art yields 69,100 results. A search for the keywords feminism and art in the Bibliography of the History of Art retrieves 1,383 results. You can even narrow Google Scholar to a subject area search of 'Physics, Astronomy and Planetary Science' and retrieve over 300 results for the keywords feminism and art. The bibliography method was preferred in this study because it yielded more manageable sets of data to compare across all four databases.

### *Evaluation Procedure*

Each art historian was tested in an author search of all four databases. All of the databases let you search by author, so articles were first retrieved by simply searching each database by author (See Figure 1). Google Scholar lacks an official authority file for author name, so some articles may have been impossible to find by using the author search.

Figure 1
Author search in 'Advanced Scholar Search'



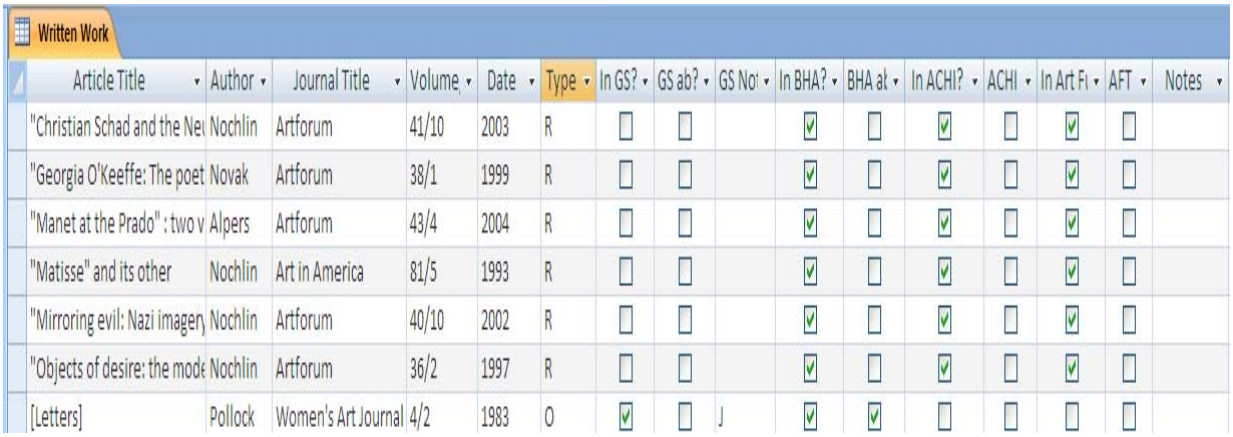

*Source:* http://scholar.google.com/advanced_scholar_search?hl=en&lr=

Limiters were set to retrieve articles:

1) Published in the year range: 1975-2008.
2) Document type set to article.
3) Language set to English-only.

All articles retrieved by author searches were recorded in an Access database. Article titles were recorded as unique identifiers. Information regarding an article's author, journal or publication title, volume/issue, date, publication type (review, article or other), were recorded in the database. All articles were identified as either being included or not included in each database (See Figure 2).

Figure 2
Screenshot of the information recorded in the Access database



Written Work

| Article Title | Author | Journal Title | Volume | Date | Type | In GS? | GS ab? | GS No | In BHA? | BHA ak | In ACHI? | ACHI | In Art Fi | AFT | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| "Christian Schad and the Ne | Nochlin | Artforum | 41/10 | 2003 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| "Georgia O'Keeffe: The poet | Novak | Artforum | 38/1 | 1999 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| "Manet at the Prado" : two v | Alpers | Artforum | 43/4 | 2004 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| "Matisse" and its other | Nochlin | Art in America | 81/5 | 1993 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| "Mirroring evil: Nazi imagery | Nochlin | Artforum | 40/10 | 2002 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| "Objects of desire: the mode | Nochlin | Artforum | 36/2 | 1997 | R | ☐ | ☐ | | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | |
| [Letters] | Pollock | Women's Art Journal | 4/2 | 1983 | O | ☑ | ☐ | J | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | |

Additionally, information was recorded as to whether or not an abstract was available for the retrieved article. An 'abstract' was considered loosely as any topical description of the record, whether it be one sentence, a whole paragraph or even a preview of the article's first page. Abstracts can be very helpful to a searcher who may not be able to discern much about the content of an article from its title and author. Additional notes were recorded for Google Scholar records regarding the source of the citation, and any publisher information. Notes were also recorded for any aberrations, such as duplicate records, in all of the records retrieved across databases.

## Results

### *Overall findings*

Google Scholar's coverage of art history literature lagged behind three art database giants with coverage of only about (35%) of the tested articles (See Table 4). Google Scholar was closely followed by Art Full Text (42%). The Arts & Humanities Citation Index covered the largest percentage of the tested articles (73%) and the Bibliography of the History of Art (BHA) trailed behind it with coverage of a little over half of the total records (54%).

Table 4
Citation, abstract and coverage of unique citation counts across four databases

| | Google Scholar | Bibliography of the History of Art | Arts & Humanities Citation Index | Art Full Text |
|---|---|---|---|---|
| Total # Citations | 163 | 257 | 344 | 196 |
| Percent Total | 35 | 54 | 73 | 42 |
| Total # Abstracts | 91 | 196 | 11 | 111 |
| Percent Abstracts | 56 | 62 | 3 | 57 |
| # Unique Citations | 18 | 44 | 90 | 22 |

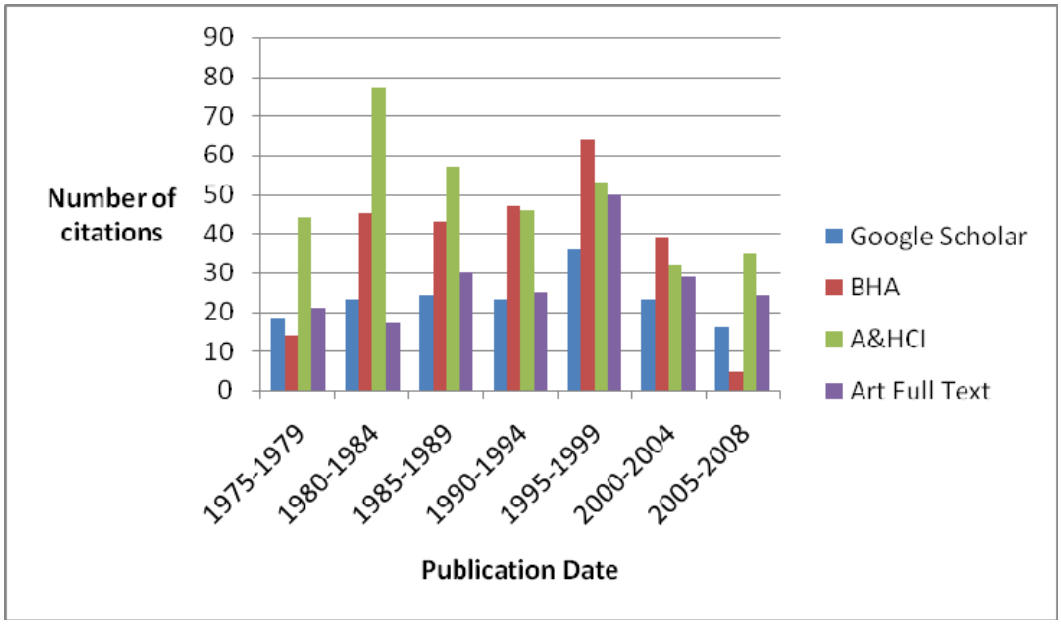
The technical definition of an 'abstract' is a concise (150- to 250 word) summary of the document (Jacsó, 2001). Abstracts can help users determine which articles to read and they can also help searchers filter results because words in an abstract will affect the relevance of the results returned. This study does not include a true measure of abstracts because it considered any synopsis or article preview as a value-added glimpse into an article that qualified it as an abstract. The BHA had the largest percentage (62%) of accompanying abstracts and the Art Index (57%) followed closely by Google Scholar (56%) (See Table 4). Abstracts in BHA adhere closely to Jacsó's definition of an abstract, whereas most of the abstracts counted in Google Scholar were previews of the full-text. The A&HCI provided a pitiful number of abstracts (3%) in comparison to the numbers provided by the other three databases, perhaps suggesting that adding abstracts is not a focus of the citation index.

Uniqueness was another factor recorded in the general findings results. Giral and Taylor's study of architecture periodical indexes pointed out the importance of database coverage analysis that considers the overlap in potentially similar subject databases. Google Scholar and Art Index contained a similar number of unique citations (18 and 22 respectively) while the BHA included 44 unique records and the A&HCI held 90 unique records (see Table 4). This study does not approach the specificity of Giral and Taylor's 1993 study on overlap because the comparison includes databases that are considered multi-disciplinary (Google Scholar and the A&CHI), but it does give a general sense of which databases have a wealth of unique citations and which do not.

*Date Coverage*

Reviewers have found inconsistencies in Google Scholar's coverage of recent publications (Neuhaus et al., 2006), in addition to its coverage of older, back-issue articles (Christianson, 2007). This study did not reveal a particular bias for Google Scholar's coverage of articles and, like Bar-Ilan's study (2006), found the wealth of back issue articles provided by JSTOR articles to contribute to its consistent coverage over the span of the four decades studied (See Figure 3). From the period of 1975-1979, Google Scholar indexed 18 citations and from the 2005-2008 period GS indexed 16 articles. Overall Google Scholar averaged an inclusion of about 23 articles per 5 year increment. The A&HCI included twice as many of the oldest citations (1975-1979) than any of the other databases. BHA does not index any of the top 15 most current articles included in the test set

Figure 3
Total number of citations by date coverage in 5 year increments across the databases




*Article type*

The article type of each citation was identified as an 'article', 'review', or 'other'. The 'other' category included letters to the editor, obituaries, poetry and other miscellaneous works published by the selected art historians. Citations identified as 'reviews' included reviews of catalogs, exhibits, works of art and reviews of monographs. Book reviews are not always considered scholarly literature, but as mentioned earlier they are significant to the discipline of art history (Garfield, 1980). While book reviews may be especially interesting to art scholars who have read the books under review, general articles may be a better fit for undergraduate students who are getting to know a subject in art history. Google Scholar included the largest percentage of articles (68%), and the smallest percentage of combined review and other types of articles (32%) (See Table 5). Despite its dominance in total citations, the A&HCI includes the smallest number of articles, but receives a boost from its superior coverage of review and other article citations.

Table 5
Citations broken down by article type (article, review, or other)

| | Google Scholar | Bibliography of the History of Art | Arts & Humanities Citation Index | Art Index/Retrospective |
|---|---|---|---|---|
| Articles | 111 | 145 | 151 | 118 |
| % Articles | 68 | 56 | 44 | 60 |
| Reviews | 31 | 81 | 141 | 54 |
| Other | 21 | 31 | 52 | 24 |
| % Reviews/Other | 32 | 44 | 56 | 40 |

*Google Scholar Sources*

A search in Google Scholar does not yield the same kind of standardized results that can be found in the other three databases. Google Scholar is, after all, a search engine, and it points users in the direction of a citation instead of providing its own citation record. Search results are retrieved from the following three sources described by Walters in his 2007 study of Google Scholar:

1) Publically accessible Web documents that "look scholarly" in content. These documents may be technical reports, dissertations, preprints and journal articles. Some are full-text, some provide citation or abstract-only access.

2) Content provided by Google Scholar's mystery partners—which includes publishers, database vendors, scholarly societies and academic institutions. Some content taken from academic institution's intranets or publishers' restricted-access sites. This is material that would not be retrieved in a regular Google search.

3) Google Scholar indexes citations found in the reference lists of the articles it finds in the above two methods. These citations do not include links to the abstract or full-text of the article. These citations are distinguishable because they include this note [citation] in front of the record. Google Scholar explains in their help

> page that these articles are ones they can't "find online." (http://scholar.google.com/intl/en/scholar/help.html) (Walters, 2007).

The articles found in this study fell into the latter two categories. By far the largest sole contributor to Google Scholar's coverage came from JSTOR (55%). Ingenta (5%), Questia (4%), Sage (4%) and a category of 13+ additional publishers identified as 'Other' (17%) (See Table 6). The second largest source of citations found in Google Scholar fall into this third category of extracted citations with 37% of the citations containing no links to abstracts or full-text (See Table 6). In addition, these citations were not complete. They lacked volume, issue and page numbers, so searchers would have to be quite persistent if they wanted to track down the citation. These were the citations that were also prone to typos and variations on author names or article titles.

Each article citation in Google Scholar refers the searcher to the website where the abstract or full-text is available from the outside source. In some cases multiple sources provide abstracts for the same citation and Google Scholar. If there is any variation or abnormality in the citation field, Google Scholar will duplicate the record. If the citation titles, authors, etc. are identical Google Scholar lists one citation with a link to the other 'versions' of the citation. A significant number (23%) of duplicates and multiple versions contributed to the total number of citations found in Google Scholar (See Table 6). While the other databases had occasional duplicates, this is an issue that is unavoidable when searching Google Scholar.

Table 6
Source of citations in Google Scholar

| Source | Citations | % Total | # Abstracts |
|---|---|---|---|
| JSTOR | 89 | 55 | 77 |
| [CITATION] | 57 | 35 | 0 |
| Ingenta | 8 | 5 | 1 |
| Questia | 7 | 4 | 6 |
| Sage | 7 | 4 | 4 |
| Other* | 27 | 17 | 19 |
| Duplicates, Triplicates, and multiple versions | 37 | 23 | N/A |

*Other publishers included: Oxford, Blackwell, U.of Chicago, dspace, Duke University Press, PEP, CSA, Intellect, Routledge, University of CA Press, Project Muse, Informaworld and one Korean publisher.

*Individual Journal Coverage*

The individual publication titles that Google Scholar indexes can only be determined by testing the database. No list or directory of journals is available to the public. Seven of the most frequently cited journals were compared across the databases to reveal any significant gaps in individual journal coverage. The most frequently cited journal, *Artforum International*, had only 4 citations appear in Google Scholar, compared to the much higher numbers available through the other three databases (See Table 7). These four articles when searched individually in Google Scholar are found to be citations extracted from the reference lists of other articles. Searching by publication type in Google Scholar for '*Artforum*' retrieves 1,830 articles, but the first one hundred articles do not provide links or abstracts suggesting that Google's crawlers do not yet

have access to articles published in *Artforum*. Searching by *Art in America* revealed similar findings. Google Scholar does index records for the *Times Literary Supplement* in general, but it did not index any of the tested art history articles. To be fair, the Art Index database does not index the *Times Literary Supplement*, *Critical Inquiry* or *Renaissance Quarterly*. The only difference is that the Art Index provides searchers with an index of what they do and do not index. From the index you find that Art Index does cover *Renaissance Quarterly*, but only from March 1967 to October 1978.

Table 7
Citation coverage in top 7 frequently cited journal publications

| | Google Scholar | Bibliography of the History of Art | Arts & Humanities Citation Index | Art Index/Retrospective |
|---|---|---|---|---|
| Artforum International (61)* | 4 | 30 | 50 | 41 |
| Art in America (52) | 16 | 34 | 46 | 42 |
| Times Literary Supplement (27) | 0 | 15 | 25 | 0 |
| Art Bulletin (24) | 19 | 13 | 22 | 14 |
| Critical Inquiry (18) | 17 | 11 | 18 | 0 |
| Art History (17) | 6 | 10 | 16 | 6 |
| Renaissance Quarterly (17) | 8 | 7 | 17 | 0 |

*The number in parentheses refers to the total number of records citing the publication.

## Discussion

While Google Scholar did not completely flunk the art history exam, it did raise some concerns that librarians should consider when deciding whether or not to recommend Google Scholar to students and scholars of art history. Neuhaus et al.

cautioned users about Google Scholar's humanities coverage which totaled only 10%, compared much more extensive coverage of science/medicine and multidisciplinary subject areas, 76% and 77% respectively (2006). This study used a bibliography method to focus specifically on art history, not under the umbrella of humanities, and found that Google Scholar covers about 35% of the kind of literature that art historians and students might require when searching for scholarly articles.

Libraries spend huge chunks of their budgets on expensive commercial databases. When a free tool with the Google brand logo arrives on the scene advertising access to the world of scholarly information, librarians must ask themselves whether certain commercial databases are worth their rising subscription costs. Librarians who have reviewed Google Scholar tend to agree that its success will depend upon "the ability of the search engine to deliver sufficient quantities of relevant and up-to-date research information" to the academic community (Neuhaus et al, 2006, p. 128). This study confirmed that, much like Google Scholar, subject-specific databases are not perfect, and they present legitimate concerns for libraries to consider when they cannot afford to subscribe to multiple art databases. Library science literature on art history has focused on the issue of images in databases and this will continue to be an important measure for how well bibliographic databases can provide scholars and students with the art history information they need. These databases should be given the same kind of scrutiny as Google Scholar, so that librarians can direct art scholars and students to the most appropriate and reliable search tool for their inquiry.

Much like Walters found in his database evaluation, Google Scholar had consistent coverage of articles by publication date. It includes articles publisher earlier

than the 1975 cutoff point for this study, which is not the case for the A&HCI (1975-present) or the Bibliography of the History of Art (1973-present). Google Scholar also includes considerable coverage of non-review articles (68%), in addition to links to the publisher's abstracts or the full text of an article. The large percentage of Google Scholar citations that are extracted from reference lists without linked abstracts (35%) raise legitimate concerns, because the citation information provided is not complete and may be confusing to a novice searcher. The large percentage of article duplicates and multiple versions (23%) also raises the question of whether Google Scholar intends to tidy up its web of scholarly citations, or whether searchers will have to sift through the massive heap of results to find relevant articles.

If a library participates in linking their resources to Google Scholar, then some of these extracted citations may be located by using the 'Get it at your library' link. Libraries can only benefit from participating in Google Scholar's 'Library Links' program because it provides patrons with one more tool for discovering academic information in the library. In an early review of Google Scholar, Carol Tenopir speculates that GS will expand user's access to library resources because it is 'intertwined with collections' (2005). Google's popularity and easy-to-use search interface will inevitably attract users and "has the potential to ensure libraries remain a critical part of the user's information-seeking process" (Pomerantz, 2006, p.55).

Google Scholar retrieves relevant results more quickly and reliably than most other commercially-produced federated search tools. It is an extremely powerful tool for searching across multiple databases for known items. By pasting just a portion of a known journal article or book into the search box a search will often return that item in

the first page of results. This may be a useful way for art historians to retrieve the literature they require for teaching and research. More contribution from art librarians and art scholars is required to determine whether literature they consider 'relevant' is indeed accessible through Google Scholar. Google Scholar claims to help searchers "identify the most relevant research" across all disciplines (see 'About Google Scholar'). Input from subject experts in art history could help determine whether this claim is accurate.

The most significant concerns among librarians and reviewers revolve around Google Scholar's lack of transparency. If Google Scholar wants to provide searchers with access to the "scholarly world of information" (See "About Google Scholar), it must also provide the public with access to the publisher and source details that define its subject scope. A journal directory, for example, would help an art historian decide whether they could expect to find articles published in the *Art Bulletin*. As of now, the only way a searcher can really know what is included in the database is to perform test searches—which probably won't fit into the busy schedules of art historians and undergraduate students.

After three-and-a-half years, Google Scholar is still a 'Beta' product. Google Scholar will shed its Beta skin at some point (presumably), and this would be an excellent opportunity for the Google team to provide the public with a more detailed description of its scope coverage. If Google Scholar released these details about what they are and where they are going, librarians would have a much easier time promoting the service as a reliable search tool in the academic library. Google Scholar has made promising strides towards partnering with academic libraries through the 'Library Links' program, but

better communication between GS and the library would help both provide better service to patrons.

## Recommendations for Further Research

The large percentage of JSTOR citations (55%) retrieved by Google Scholar could warrant a study in itself. JSTOR, a digital archive of scholarly publications, indexes journal content dating back to 1665 (See http://www.jstor.org/) and it would be informative to examine how consistently Google Scholar's crawls and indexes the JSTOR archive. Research in the arts and humanities often relies heavily on back-issue journals, so a study focusing on this type of coverage would be useful.

Monograph and book publication in art history is as important, if not more so, than publications in the periodical literature. Both Google Scholar and the Bibliography of the History of Art index citations to monographs, as well as articles, and it would be interesting to see how well Google Scholar indexes these publications. From test searches it is evident that Google Scholar has access to OCLC Worldcat holdings (Tenopir, 2005), but more studies are needed to determine the extent of GS's coverage of these holdings.

Further content evaluation of Google Scholar in the arts and humanities will help to determine the scholarly search engine's strengths and weaknesses. As stated previously, Google Scholar is an attractive tool for undergraduates and librarians must be prepared to discuss its strengths and weaknesses with students at the reference desk or in the library instruction classroom. This study approached Google Scholar from the perspective of a university student or scholar who would have access to many of the subscription databases available through an academic library. Google Scholar is a free

service and continued studies targeting the open access literature that GS offers would help searchers who are not affiliated with a university get a sense of the tool's usefulness for their information needs.

## Conclusion

This study reveals that Google Scholar hasn't yet achieved its goal to provide searchers with complete access to the most relevant research in art history. Jacsó calls upon librarians and professional searchers to act as the "final control team" to corroborate what database producers say and what databases actually include (2001, p.8). This study responded to this call by evaluating content coverage across databases on subject-specific literature. By dissecting Google Scholar's art history coverage, this study revealed scope details pertaining to date, source and abstract coverage. Additionally, this study picked individual publications in the test set to analyze and discuss previously unknown information about the sources included, or not included, in Google Scholar. By examining Google Scholar through the lens of Jacsó's scope criteria, a clearer picture emerged as to what a seeker of art history literature can expect to find when searching the scholarly search engine.

# Acknowledgements

I would like to thank the following people for their support in the writing of this paper

▪ My advisor, Jeffrey Pomerantz, for his sage advice and suggestions.

▪ My colleague, Lee Sorenson, Art Librarian at Duke University and author of the *Dictionary for Art Historians*, for his interest in the paper and his willingness to share his expert knowledge in all things art historical.

▪ My favorite database critic, Péter Jacsó, for the intelligence and humor he brings to evaluating databases.

▪ My mother, who kindly lent her encouragement and editing expertise from faraway Uganda.

# References

"About Google Scholar." http://scholar.google.com/intl/en/scholar/about.html

"Google Scholar Help." http://scholar.google.com/intl/en/scholar/help.html

Abram, S. (2005). Google Scholar: thin edge of the wedge? *Information Outlook*, *9*(1), 44-46.

Adlington, J. & Benda, C. (2006). Checking under the hood: evaluating Google Scholar for reference use. *Internet Reference Services Quarterly*, *10*(3/4), 135-148. doi: 10.1300/J136v10n03_12

Bar-Ilan, J. (2006). An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing and Management*, *42*, 1553-1566. doi: 10.1016/j.ipm.2006.03.019

Beaudoin, J. (2005). Image and text: A review of the literature concerning the information needs and research behaviors of Art Historians.*Art Documentation, 24*(2), 34-37. Retrieved from http://www.hwwilson.com/

Burright, M. (2006). Database reviews and reports: Google Scholar—Science & Technology. *Issues in Science and Technology Librarianship, 45*. Retrieved March 18, 2008, from http://www.istl.org/06-winter/databases2.html.

Callicott, B. & Vaughn, D. (2005). Google Scholar vs. Library Scholar: testing the performance of Schoogle. *Internet Reference Services Quarterly, 10*(3/4), 71-88. doi: 10.1300/J136v10n03_08

Christianson, M. (2007). Ecology articles in Google Scholar: Levels of access to articles in core journals. *Issues in Science and Technology Librarianship*, *49*. Retrieved February 24, 2008, from http://www.istl.org/07-winter/refereed.html .

deLuise, A. (2003). Full text or not? All illustrations or not? *Art Documentation, 22*(2), 20-25. Retrieved from http://www.hwwilson.com/

Gardner, S. & Eng S. (2005). Gaga Over Google? Scholar in the Social Sciences. *Library Hi Tech News, 22*(8), 42-45. doi: 10.1108/07419050510533952

Garfield, E. (1980). Is information retrieval in the arts and humanities inherently different from that in science? *Library Quarterly*, *50*(1), 40-57.

Giral, A. & Taylor, A.G. (1993). Indexing overlap and consistency between the Avery Index to Architectural Periodicals and the Architectural Periodicals Index. *Library Resources and Technical Services*, *37*(1), 19-43.

Hughes, T. (Interviewer) & Acharya, A. (Interviewee). (2006). An interview with Anurag Acharya, Google Scholar lead engineer. Retrieved April 1, 2008, from Google 'Librarian Central': http://www.google.com/librariancenter/articles/0612_01.html.

Jacsó, P. (2005a). As we may search—Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced Databases. *Current Science*, *89*(9), 1537-1547. Retrieved May 28, 2007, from Directory of Open Access Journals.

Jacsó, P. (1997). Content evaluation of databases. *Annual Review of Information Science and Technology (ARIST)*, *32*, 231-267.

Jacsó, P. (2001). *Content evaluation of textual CD-ROM and web databases*. Libraries Unlimited: Englewood, CO.

Jacsó, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review*, *30*(3): 297-309. Retrieved from http://www.emeraldinsight.com

Jacsó, P. (2005b). Google Scholar: The pros and cons. *Online Information Review*, *29*(2), 208-214. Retrieved from http://www.emeraldinsight.com

Jacsó, P. (2005c). Péter's Digital Reference Shelf: Google Scholar (redux). Retrieved April 1, 2008, from Gale CENGAGE learning: http://www2.hawaii.edu/~jacso/.

Jacsó, P. (2008). Google Scholar revisited. *Online Information Review, 32*(1), 102-114. Retrieved from http://www.emeraldinsight.com

Kesselman, M. & Watstein, S.B. (2005). Google Scholar and libraries: Point/counterpoint. *Reference Services Review*, *33*(4), 380- 387. doi: 10.1108/00907320510631535

Lancaster, F.W. (1971). The evaluation of published indexes and abstract journals. *Bulletin of the Medical Library Association*, *59*, 479-494. Retrieved February 25, 2008, from PubMed Central.

Martyn J. & Slater, M. (1964). Tests on abstracts journals: Coverage, overlap and indexing. *Journal of Documentation, 20*(4), 212-235.

Mayr, P. & Walter, A. (2007). An exploratory study of Google Scholar. *Online Information Review, 31*(6), 814-830. Retrieved from http://www.emeraldinsight.com

Neuhaus, C., et al. (2006). The Depth and breadth of Google Scholar: An empirical study. *portal: Libraries and the Academy, 6*(2), 127-141. Retrieved from http://muse.jhu.edu/journals/portal_libraries_and_the_academy/

Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *Libri, 55*(4), 170-180. Retrieved January 11, 2008, from Miscellaneous E-Journals database: http://www.librijournal.org/2005-4toc.html

Pomerantz, J. (2006). Google Scholar and 100 percent availability of information. *Information Technology and Libraries*. *25*(2), 52-56.

Quint, B. (2007). Changes at Google Scholar: A conversation with Anurag Acharya. *Information Today, Inc.* Retrieved April 1, 2008, from http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=37309

Schroeder, R. (2007). Pointing users toward citation searching: Using Google Scholar and Web of Science. *portal: Libraries and the Academy*, *7*(2), 243-248. Retrieved from http://muse.jhu.edu/journals/portal_libraries_and_the_academy

Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association*, *95*(4), 442-453. doi: 10.3163/1536-5050.95.4.442

Sorensen, L. (ed.) (2006). *Dictionary of Art Historians*. http://www.dictionaryofarthistorians.org/

Tenopir, C. (1982). Evaluation of database coverage: A comparison of two methodologies. *Online Review*, *6*(5), 423-441.

Tenopir, C. (2005). Online databases: Google in the academic library. *Library Journal*, *130*(2), 32. Retrieved May 23, 2007, from H.W.Wilson Web.

Walters, W.H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, *43*(4), 1121-1132. doi: 10.1016/j.ipm.2006.08.006

White, B. (2006). Examining the claims of Google Scholar as a serious information source. *New Zealand Library & Information Management Journal*, *50*(1), 11-24. Retrieved March 17, 2008, from http://eprints.rclis.org/7657.

Wlekinski, J.M. (2005). Studying Google Scholar: Wall to wall coverage? *Online*, *29*(3), 22-26.

# Appendix

| | Google Scholar | Bibliography of the History of Art (BHA) | Art & Humanities Citation Index (A&CHI) | Art Full Text/ Art Index Retrospective |
|---|---|---|---|---|
| **Language** | Unknown; Increasingly International (see Literature Review for more details) | International | English | French, Italian, German, Japanese, Spanish, Dutch and Swedish |
| **Dates Included** | Unknown; test searches show results for 20th century to present | 1973 to present | 1975 to present | 1929-1984 (Art Retrospective); 1984 to present (AFT) |
| **Currency** | Unknown | Every 3 months | Weekly | Updated daily |
| **Topical coverage** | Multidisciplinary, 'all broad areas of research' | European and American visual arts; fine arts as well as decorative and applied arts | Multidisciplinary; journal literature of the arts and humanities | Advertising art, antiques, archaeology, architecture/architectural history, art history, non-Western art, painting, sculpture, and more. |
| **Size** | Unknown | 650, 824 + records; 4,819+ journals | 2.5 million records; 1,160 arts & humanities journals | 480 + periodicals |
| **Document Types** | Peer-reviewed journals, papers, theses, books | Journals, art-related books, conference proceedings, dissertations, exhibition and dealers' catalogs | Journals | Periodicals, yearbooks and museum bulletins |

*Sources: From Help/Information pages provided within each database.*