# Comparing test searches in PubMed and Google Scholar **EC**

**Mary Shultz, MS, AHIP**

See end of article for author's affiliations.

## INTRODUCTION

Google Scholar has been met with both enthusiasm and criticism since its introduction in 2004. This search engine provides a simple way to access ''peer-reviewed papers, theses, books, abstracts, and articles from academic publishers' sites, professional societies, preprint repositories, universities and other scholarly organizations'' [1]. An obvious strength of Google Scholar is its intuitive interface, as the main search engine interface consists of a simple query box. In contrast, databases, such as PubMed, utilize search interfaces that offer a greater variety of advanced features. These additional features, while powerful, often lead to a complexity that may require a substantial investment of time to master. It has been observed that Google Scholar may allow searchers to ''find some resources they can use rather than be frustrated by a database's search screen'' [2]. Some even feel that ''Google Scholar's simplicity may eventually consume PubMed'' [3].

Along with ease of use, Google Scholar carries the familiar ''Google'' brand name. As Kennedy and Price so aptly stated, ''College students AND professors might not know that library databases exist, but they sure know Google'' [4]. The familiarity of Google may allow librarians and educators to ease students into the scholarly searching process by starting with Google Scholar and eventually moving to more complex systems. Felter noted that ''as researchers work with Google Scholar and reach limitations of searching capabilities and options, they may become more receptive to other products'' [5].

Google Scholar is also thought to provide increased access to gray literature [2], as it retrieves more than journal articles and includes preprint archives, conference proceedings, and institutional repositories [6]. Google Scholar also includes links to the online collections of some academic libraries. Including these access points in Google Scholar retrieval sets may ultimately help more users reach more of their own institution's subscriptions [7].

While its advantages are substantial, Google Scholar is not without flaws. The shortcomings of the system and its search interface have been well documented in the literature and include lack of reliable advanced search functions, lack of controlled vocabulary, and issues regarding scope of coverage and currency. Table 1 summarizes some of the reported criticisms of Google Scholar.

Vine found that while Google Scholar pulls in data from PubMed, many PubMed records are missing

**EC** Supplemental Table 5 and an appendix are available with the online version of this journal.

[20], and that Google Scholar also lacks features available in MEDLINE [12]. Others have noted that Google Scholar should not be the first or sole choice when searching for patient care information, clinical trials, or literature reviews [23, 24]. Thorough review and testing of Google Scholar, being an approach similar to that used to evaluate licensed resources, is necessary to better understand its strengths and limitations. As Jacso states, ''professional searchers must do sample test searches and correctly interpret the results to corroborate claims and get factual information about databases'' [18]. This paper compares and contrasts a variety of test searches in PubMed and Google Scholar to gain a better understanding of Google Scholar's searching capabilities.

## METHODOLOGY

Ten searches were performed in PubMed using a variety of available search features. The searches were repeated in Google Scholar to approximate a user's approach to those same topics in that search engine. The searches, performed between August and September 2006, were by topic, author, title, journal name, and/or combinations of those fields (Appendix online). Topics included iron-deficiency anemia, bupropion for smoking cessation, and articles by specific authors in specific journals. The topics selected were loosely based on questions received during reference transactions or were previously developed for use during instruction.

For each search, the citations received via Google Scholar and PubMed were examined to determine a variety of characteristics including format, date, Medical Subject Headings (MeSH) where appropriate, uniqueness, duplications, and full-text availability from the author's institution.

Most searches were narrowed by date to produce sets of a reasonable size to allow comparison of unique items retrieved by each system. The search results were analyzed to determine possible reasons for the retrieval of unique items in each resource and to gather information on the general features of the Google Scholar results.

## RESULTS

In eight of the ten searches, Google Scholar returned larger retrieval sets than PubMed (Table 2). Table 3 illustrates the characteristics of the items retrieved by Google Scholar, and Table 4 provides information on PubMed retrieval sets. Most items retrieved by Google Scholar were journal articles (Table 3). Items in other formats included: 9 books, 11 book reviews, 2 Web pages, 1 subject index listing, 1 thesis, 1 newsletter item, 1 bibliography, 4 author replies, 1 annual meeting abstract, and 1 draft document. These results yielded few gray literature items.

The main title link in Google Scholar citations was used to determine if full text was found. Full text was available in 46.96% (116/247) of the total citations retrieved. In most cases, it was assumed that full-text access was based on the institutional subscriptions

**Table 1**
Criticisms of Google Scholar

| Criticisms | References |
| --- | --- |
| Advanced search functions may be unreliable | [8–10] |
| No ability to search controlled vocabulary or no authority control for journal names or author names | [10–13] |
| Some materials retrieved may not be scholarly | [14] |
| Secretive about how it defines "scholarly" | [15] |
| Secretive about scope or coverage | [5, 9, 10, 12, 13, 16–18] |
| May not be current | [9–11, 14, 19] |
| Missing PubMed records | [11, 20] |
| Lack of sorting options* | [10, 11, 14] |
| Inclusion of duplicate citations in results | [6, 14] |
| Only the first 1,000 results can be viewed | [4, 19, 21] |
| Not as comprehensive or precise as searching native interfaces | [9, 11, 22] |
| Lack of limiting features | [11, 12] |

* In the spring of 2006, Google Scholar introduced an option to re-sort with more current citations appearing first.

available to the author of this study. Some items retrieved might have been freely available. In 22.67% (56/247) of the results, the Google Scholar citation was simply a link out to a PubMed record. As shown in Table 4, nearly half (48.98%; 72/147) of PubMed citations provided full-text access through the author's institution.

The unique items retrieved by each interface were examined to determine why they were missed by the other system. Across all searches, Google Scholar retrieved a total of 247 citations, 125 (50.61%) of which were unique to Google Scholar. Analysis revealed the following characteristics:

■ Thirty-two items (12.96%) retrieved by Google Scholar were formats other than journal articles.
■ Some unique Google Scholar items (10 items, 4.05%) appeared in journals not indexed by PubMed.
■ Google Scholar covered a wider date range and returned 4 items (1.62%) older than 1950 that were not in PubMed.
■ Google Scholar retrieved items based on its ability to search the full text of many articles rather than solely on citation data.

PubMed retrieved a total of 147 citations across all searches, and, of these, 46 (31.29%) were unique.

## DISCUSSION

Assumptions of search engine performance based purely on retrieval quantities can be misleading without closer investigation of the results. For example, Table 2 shows that many of the searches returned quantities that were close in numbers. In search #1 (dietary supplements as a treatment for iron deficiency anemia), PubMed returned twenty-five citations, while Google Scholar returned twenty-six citations. However, only four citations were common to both systems. In search #2 (Mobius syndrome), Google Scholar returned eleven citations, while PubMed found ten citations but with an overlap of only two citations retrieved by both systems.

Terminology was observed to be a major factor affecting retrieval and the ability of both systems to return unique items. Some unique items retrieved by Google Scholar were off topic. These "false hits" appear to be related to Google Scholar's full-text searching along with a lack of controlled vocabulary. For example, the purpose of search #7 was to find articles on the topic of "wine" that appeared in the *New England Journal of Medicine.* Google Scholar retrieved eight items where the word "wine" appeared in the full text but was not the main topic of the article, in one case, retrieving an article where the authors acknowledge a colleague with the surname Wine. Google Scholar also returned items that contained the search terminology but did not match the intention of the search. In the search for information about dietary supplements in the treatment of iron deficiency (search #1), Google Scholar returned some citations about high iron stores rather than deficiency (Table 5 online). Google Scholar searches for a word or sequence of letters and not the concept or meaning.

The complete citations for all unique items retrieved by PubMed were examined. One possible explanation why Google Scholar failed to retrieve the same items was that many were indexed under the appropriate MeSH term, although the search phrase might not have appeared in the title or abstract. For example, search #9 was designed to retrieve articles by Visek about the topic of ammonia. While ammonia was not searched specifically as a MeSH term, PubMed automatically mapped it to MeSH. Of the unique citations retrieved by PubMed, some were indexed under ammonia although this term did not appear in the citation (Table 5 online). While Google Scholar offers the ability to use a tilde (~) to retrieve alternative termi-

**Table 2**
Number of retrieved items

| Search # | PubMed results | PubMed unique items | Google Scholar results | Google Scholar unique items |
| --- | --- | --- | --- | --- |
| 1 | 25 | 21 | 26 | 20 |
| 2 | 10 | 8 | 11 | 8 |
| 3 | 4 | 2 | 27 | 24 |
| 4 | 10 | 3 | 52 | 38 |
| 5 | 6 | 0 | 20 | 10 |
| 6 | 11 | 0 | 20 | 8 |
| 7 | 2 | 1 | 10 | 8 |
| 8 | 13 | 0 | 18 | 4 |
| 9 | 51 | 7 | 49 | 4 |
| 10 | 15 | 4 | 14 | 1 |

**Table 3**
Characteristics of Google Scholar results

| Google Scholar | Search numbers | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 |
| Total number of results | 26 | 11 | 27 | 52 | 20 | 20 | 10 | 18 | 49 | 14 |
| A. Journal article citations | 23 | 11 | 22 | 42 | 17 | 12 | 9 | 18 | 48 | 13 |
| B. Book citations | 3 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 1 | 0 |
| C. Book reviews | 0 | 0 | 0 | 5 | 2 | 4 | 0 | 0 | 0 | 0 |
| D. Other | 0 | 0 | 2 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| E. Older than 1950 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| F. Duplicate in Google Scholar set | 0 | 0 | 0 | 8 | 5 | 1 | 1 | 1 | 1 | 2 |
| G. Also in PubMed set | 4 | 2 | 2 | 13 | 10 | 12 | 2 | 14 | 45 | 13 |
| H. Link to a PubMed record | 4 | 0 | 2 | 7 | 1 | 0 | 0 | 9 | 31 | 2 |
| I. If unique, item found in PubMed directly | 16 | 8 | 17 | 24 | 9 | 0 | 7 | 0 | 2 | 0 |
| J. Page not found or error | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| K. Google Scholar item title linked to full text | 13 | 7 | 18 | 34 | 6 | 19 | 5 | 2 | 3 | 9 |
| Total number of unique Google Scholar items | 20 | 8 | 24 | 38 | 10 | 8 | 8 | 4 | 4 | 1 |

Note that item G will sometimes contain numbers greater than the retrieval set of PubMed. This occurred in cases in which Google Scholar returned duplicate citations that matched a single citation in PubMed. See online Appendix for search strategy details.

nology, this ability does not provide the control that subject headings do.

## CONCLUSION

Performing a direct and exact comparison between searches in Google Scholar and PubMed is not possible as the systems function in very different manners. For example, PubMed searches a well-defined set of journals, while Google Scholar includes resources beyond journals and the exact scope of coverage is not extensively described. Because the systems are not searching identical data, the results are often different.

Although these two systems are difficult to compare, it is still important to explore the differences between them. Librarians should understand the strengths and weaknesses of Google Scholar and be prepared to explain them to their users [14]. It may also be wise to consider including Google Scholar in bibliographic instructional sessions and to convey how it compares to other search interfaces [11]. For example, Google Scholar does not offer the number and extent of special searching and limiting features available in PubMed. However, Google Scholar provides some advantages in that it is an easy place to begin a search to find an initial retrieval of possibly worthwhile articles. It also offers searchers the ability to find citations to older items that they would miss if they use only PubMed. Additionally, Google Scholar has the potential to provide access to the gray literature. This increased access to a part of the biomedical literature, which can be difficult to search, may have implications for the public health field [25].

One of the most advantageous features of searching PubMed is the ability to utilize the MeSH vocabulary, as Google Scholar does not currently implement controlled vocabulary searching mechanisms. MeSH provides a powerful method of narrowing results and homing in on what the searcher needs. PubMed also offers substantially more features that allow searchers to narrow their retrieval to citations from clearly identified sources, as detailed in NLM's List of Journals Indexed for MEDLINE and List of Serials Indexed for Online Users [26]. The problem faced today by searchers is not a lack of information but rather an overload of information. For a researcher conducting human studies, writing a dissertation, finding information pertinent to patient care, or conducting an in-depth literature review, Google Scholar does not appear to be a replacement for PubMed, though it may serve effectively as an adjunct resource to complement databases with more fully developed searching features. It is important to note that both PubMed and Google Scholar are often upgraded with new features or with intended improvement of existing functions. It may be worthwhile to repeat this study in one or two years to determine if further refinements have improved their performance.

## ACKNOWLEDGMENTS

**Table 4**
Characteristics of PubMed results

| PubMed | Search numbers | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 |
| Total number of results | 25 | 10 | 4 | 10 | 6 | 11 | 2 | 13 | 51 | 15 |
| Also in Google Scholar set | 4 | 2 | 2 | 7 | 6 | 11 | 1 | 13 | 44 | 11 |
| Item provides library link out to full text* | 19 | 7 | 2 | 6 | 3 | 11 | 2 | 4 | 4 | 14 |
| Total number of unique PubMed items | 21 | 8 | 2 | 3 | 0 | 0 | 1 | 0 | 7 | 4 |

* Library link out icon was for author's institution. See online Appendix for search strategy details.

Beck, Sandra De Groote, AHIP, Victoria Pifalo, and Ann Carol Weller.

## REFERENCES

1. Google Scholar help: about Google Scholar. [Web document]. Mountain View, CA: Google, 2005 [cited Aug 2006]. <http://scholar.google.com/intl/en/scholar/about.html>.
2. Kesselman M, Watstein SB. Google Scholar(tm) and libraries: point/counterpoint. Ref Serv Rev 2005; 33(4):380–7.
3. Abbasi K. Simplicity and complexity in health care: what medicine can learn from Google and iPod. J R Soc Med 2005 Sep;98(9):389.
4. Kennedy S, Price G. Big news: ''Google Scholar'' is born. [Web document]. Resource Shelf. UK: Free Pint Limited, 2004. [cited Aug 2006]. <http://www.resourceshelf.com/2004/11/18/wow-its-google-scholar/>.
5. Felter LM. Google Scholar, Scirus, and the scholarly search revolution. Searcher 2005 Feb;13(2):43–8.
6. Giles J. Science in the Web age: start your engines. Nature 2005 Dec 1;438(7068):554–5.
7. Notess GR. Scholarly Web searching: Google Scholar and Scirus. Online 2005 Jul–Aug;29(4):39–41.
8. Jacso P. Google Scholar (redux). [Web document]. Peter's Digital Reference Shelf, 2005. [cited Aug 2006]. <http://www.galegroup.com/reference/archive/200506/google>.
9. Jacso P. Google Scholar: the pros and the cons. Online Inform Rev 2005;29(2):208–14.
10. Burright M. Google Scholar: science & technology. [Web document]. Issues Sc Technol Libr 2006 Winter;45. [cited Aug 2006]. <http://www.istl.org/06-winter/databases2.html>.
11. Giustini D, Barsky E. A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations. J CHLA/J ABSC 2005;26:85–9.
12. Vine R. Google Scholar [electronic resources review]. J Med Libr Assoc 2006 Jan;94(1):97–9.
13. Myhill M. The ADVISOR reviews . . . Google Scholar. [Web document]. The Charleston ADVISOR 2005 Apr;6(4). [cited Sep 2006]. <http://www.charlestonco.com/review.cfm?id=225>.
14. Gardner S, Eng S. Gaga over Google? Scholar in the social sciences. Library Hi Tech News 2005;8:42–5.
15. Friend FJ. Google Scholar: potentially good for users of academic information. J Electronic Publishing [serial online]. 2006 Jan;9(1). [cited Sep 2006]. <http://hdl.handle.net/2027/spo.3336451.0009.105>.
16. Steinbrook R. Searching for the right search: reaching the medical literature. N Engl J Med 2006 Jan 5;354(1):4–7.
17. Wleklinski JM. Studying Google Scholar: wall to wall coverage? Online 2005 May–Jun;29(3):22–6.
18. Jacso P. 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 2005 Nov;89(9):1537–47.
19. Tennant R. Google, the naked emperor. Libr J 2005 Aug; 130(13):29.
20. Vine R. Google Scholar is a full year late indexing PubMed content. [Web document]. SiteLines, Feb 2005. [cited Aug 2006]. <http://www.workingfaster.com/sitelines/archives/2005_02.html#000282>.
21. Crawford W. Google: a company, not a religion. EContent 2005 Nov;28(11):42.
22. Jacso P. Side-by-side2, native search engines vs. Google Scholar. [Web document]. University of Hawai'i, 2005. [cited Aug 2006]. <http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm>.
23. Henderson J. Google Scholar: a source for clinicians? CMAJ 2005 Jun 7;172(12):1549–50.
24. Giustini D. How Google is changing medicine. BMJ 2005 Dec 24;331(7531):1487–8.
25. Turner AM, Liddy ED, Bradley J, Wheatley JA. Modeling public health interventions for improved access to the gray literature. J Med Libr Assoc 2005 Oct;93(4):487–94.
26. National Library of Medicine. List of journals indexed for MEDLINE and list of serials indexed for online users terms and conditions. [Web document]. Bethesda, MD: The Library, 2007. [cited 29 Mar 2007]. <http://www.nlm.nih.gov/tsd/serials/terms_cond.html>.

## AUTHOR'S AFFILIATION

**Mary Shultz, MS, AHIP,** shultz@uic.edu, Assistant Health Sciences Librarian, Library of the Health Sciences–Urbana, University of Illinois at Chicago, 102 Medical Sciences Building, 506 South Mathews Avenue, Urbana, IL 61801

# Characteristics of cancer blog users ▉EC

**Sujin Kim, PhD; Deborah S. Chung, PhD**

See end of article for authors' affiliations.

## INTRODUCTION

Blogs are a relatively new medium in computer-mediated health communication and are regarded as highly opinionated journals maintained by millions of users who read and write personal remarks on issues ranging from news stories to health care [1–3]. Of the 120 million US adults with Internet access, 7%, or 8 million people, have created blogs [4], and the increasing use of blogs has been reported in several studies [1, 4, 5]. Rainie found that the typical blogger is a young, male, Internet veteran; has a broadband connection; and is financially secure [5]. The gender of the blogger has also been a topic for research. Herring et al. found that even though women participate in blogging activities (focusing on emotional support), men are more likely to create filter blogs and k-logs (knowledge blogs) that are considered focused on information [1].

Blogs have been described as a new medium, one that shifts mainstream control of information into the hands of the audience. The potential use of blogs for cancer patients, basic scientists, clinical researchers, and practicing oncologists to discuss findings and suggestions has been envisioned in several cancer journals [6]. In addition, the use of online communication tools to share emotional support in all aspects of cancer-related issues has been frequently described [2, 6].

▉EC Figures 1 and 2 and a supplemental appendix are available with the online version of this journal.

While blogs are becoming more frequently researched, empirical studies regarding blogs and their users, especially cancer patients and their companions (defined for this study as patient family and friends), are noticeably lacking. Most research has been in the area of news media [7]. Some research has been reported in lexical (textual) analyses from studies designed to provide technological frameworks to classify blog messages for improved accessibility [8, 9]. However, questions regarding motivation to post or comment on blogs and the perceived outcomes of using blogs still remain challenging research tasks [10, 11]. Understanding how blogs are used can allow information providers to better understand the impact blogs can have on cancer patients, their friends, and families. This study used cluster analysis techniques to classify cancer blog users' demographics, as well as their use and perceptions of blogs.

## METHODS

The study, approved by the University of Kentucky's Institutional Review Board, used an online survey to target users of blogs with cancer-related content. Invitations to participate in the survey were posted on 153 individual personal blogs* that were identified through bi-weekly searches by the authors between March 30 and June 3, 2006, using Google Blog Search [12]. Searches were limited to include only blogs with the word ''cancer'' in their titles, written in the English language, and with posts created within the past month.* In addition, the search was limited to frequently used blogging services, such as Blogspot, LiveJournal, and Typepad. Before participating in the survey, individuals read a study information sheet and provided consent to participate in the study.

Survey questions sought demographic information, usage, motivation, behavioral changes, and limitations of using blogs among cancer blog users (Appendix online). The survey questions were designed and modified based on previous cancer research and research on motivations for using the Internet [13–15]. In particular, questions about motivations for blog use were modified to better encompass cancer patients, family, and friends as prior research typologies focused on entertainment, diversion, and habitual motivations and were not appropriate or relevant for this study. In completing the survey, participants were asked to identify themselves as a cancer patient, companion, or health care provider.

Cluster analysis can be useful in identifying natural groupings of homogeneous groups of people in a manner that both minimizes within-group variations and maximizes between-group variations [16–22]. Because little is known about cancer blog users, this study used cluster analysis to classify users and ascertain patterns of characteristics represented by those groups. The optimal number of clusters was determined based on the visual analysis of clustering results, such as a dendro-gram and scree diagram. The significance of the clustering results was based on the Best-Cut suggestion by Mojena rules 1 and 2.

## RESULTS

### Overall sample demographics

The survey was completed by 113 respondents; 59.29% (n = 67) were cancer patients; 31.86% (n = 36) were friends or family of cancer patients; and 6.19% (n = 7) were health care providers. Three participants did not answer this question. About 77% of respondents were female (n = 87), and 22.12% were male (n = 25). About 94% percent (n = 99) of the sample were Caucasian. The most frequently reported salary earned by the population was between $60,000 and $75,000. The average age of the respondents was 57, and 91 (71.68%) respondents held bachelor's degrees or higher.

### Characteristics of cancer bloggers

Cluster analysis revealed three clusters among the data. Visual analyses of the dendrogram and scree diagram (Figures 1 and 2, online) confirmed the three-cluster solution as the optimal number for the given data set. Table 1 displays each cluster group's general demographics, and Table 2 displays the means of the variables blog usage, motivation, behavioral changes, and limitations of blog use.

### Demographic characteristics

Cluster 1 included 38 (33.63%) bloggers whose average age was slightly higher than that of the sample (40.82, SD 11.03, vs. 40.26, SD 12.31). Among the 38 members in cluster 1, 25 were patients and 9 were friends or family members. Over 71% (n = 27) of cluster 1 members answered that they hosted their own blogs, again representing the highest among the 3 clusters. Cluster 2 had the most members in its group (n = 48, 42.48%). Similar to the other 2 clusters, Caucasian women dominated this group (n = 33, 68.75%). Eighteen of 30 bloggers in this cluster were single. The number of friends and family members found in this group was 20, more than half of the total friends and family members in this study. Cluster 3 had the fewest male bloggers (n = 4, 16%), and the average age of the members was 40.63 (SD 11.96). Cluster 3 also included the least number of individuals who hosted their own blogs (n = 15, 55.55%).

### Blog use

Cluster 1 used blogs for an average of 16.76 months, which was slightly less than the other 2 clusters. Cluster 2 reported higher mean scores (4.11) for seeking health care providers as their information source than the other 2 clusters. Information sought in medical libraries and patient education centers was more frequently sought in cluster 3 (mean score 2.92) compared to the other 2 clusters.

---

* The invitation was only posted in personal blogs that do not restrict postings and comments by the public.

**Table 1**
General characteristics of the three identified clusters

| General characteristics | Cluster 1 (n = 38, 33.63%) n (%) | Cluster 2 (n = 48, 42.48%) n (%) | Cluster 3 (n = 27, 23.89%) n (%) | Total (n = 113, 100%) n (%) |
|---|---|---|---|---|
| **Age** | 40.82 (SD: 11.03) | 39.6 (SD: 12.31) | 40.63 (SD: 12.95) | 40.26 (SD: 11.96) |
| **Gender** | | | | |
| Female | 26 (68.42) | 38 (79.17) | 23 (85.19) | 87 (76.99) |
| Male | 11 (28.95) | 10 (20.83) | 4 (14.81) | 25 (22.12) |
| Missing | 1 (2.63) | 0 (—) | 0 (—) | 1 (0.88) |
| **Education** | | | | |
| High school | 7 (18.42) | 5 (10.42) | 4 (14.81) | 16 (14.16) |
| Associate | 4 (10.53) | 10 (20.83) | 2 (7.41) | 16 (14.16) |
| Bachelor | 11 (28.95) | 21 (43.75) | 12 (44.44) | 44 (38.94) |
| Master | 10 (26.32) | 8 (16.67) | 5 (18.52) | 23 (20.35) |
| Doctorate | 6 (15.79) | 4 (8.33) | 4 (14.81) | 14 (12.39) |
| **Income** | | | | |
| a. $10,000–$19,999 | 1 (2.63) | 6 (12.50) | 0 (—) | 7 (6.19) |
| b. $20,000–$29,999 | 3 (7.89) | 4 (8.33) | 3 (11.11) | 10 (8.85) |
| c. $30,000–$39,999 | 3 (7.89) | 8 (16.67) | 2 (7.41) | 13 (11.50) |
| d. $40,000–$49,999 | 6 (15.79) | 1 (2.08) | 0 (—) | 7 (6.19) |
| e. $50,000–$59,999 | 1 (2.63) | 3 (6.25) | 6 (22.22) | 10 (8.85) |
| f. $60,000–$75,000 | 5 (13.16) | 10 (20.83) | 1 (3.70) | 16 (14.16) |
| g. $75,001–$99,999 | 7 (18.42) | 2 (4.17) | 5 (18.52) | 14 (12.39) |
| h. $100,000–$150,000 | 2 (5.26) | 3 (6.25) | 5 (18.52) | 10 (8.85) |
| i. $151,000–$199,999 | 1 (2.63) | 1 (2.08) | 0 (—) | 2 (1.77) |
| j. $200,000–$250,000 | 0 (—) | 2 (4.17) | 0 (—) | 2 (1.77) |
| k. $250,000+ | 0 (—) | 1 (2.08) | 1 (3.70) | 2 (1.77) |
| Missing | 9 (23.68) | 7 (14.58) | 4 (14.81) | 20 (17.70) |
| **Ethnicity** | | | | |
| Caucasian | 30 (78.95) | 42 (87.50) | 27 (1.00) | 99 (87.61) |
| African American | 2 (5.26) | 0 (—) | 0 (—) | 2 (1.77) |
| Asian American | 1 (2.63) | 0 (—) | 0 (—) | 1 (0.88) |
| Native American | 0 (—) | 1 (2.08) | 0 (—) | 1 (0.88) |
| Hispanic | 0 (—) | 2 (4.17) | 0 (—) | 2 (1.77) |
| Missing | 5 (13.16) | 3 (6.25) | 0 (—) | 8 (7.08) |
| **Marital status** | | | | |
| Single | 9 (23.68) | 18 (37.50) | 3 (11.11) | 30 (26.55) |
| Married | 19 (50.00) | 22 (45.83) | 20 (74.07) | 61 (53.98) |
| Partnered | 2 (5.26) | 2 (4.17) | 1 (3.70) | 5 (4.42) |
| Divorced | 6 (15.79) | 4 (8.33) | 3 (11.11) | 13 (1.15) |
| Widowed | 0 (—) | 2 (4.17) | 0 (—) | 2 (1.77) |
| Refused | 1 (2.63) | 0 (—) | 0 (—) | 1 (0.88) |
| Missing | 1 (2.63) | 0 (—) | 0 (—) | 1 (0.88) |
| **Providers, patients, companions** | | | | |
| Providers | 3 (7.89) | 3 (6.25) | 1 (3.70) | 7 (6.19) |
| Patients | 25 (65.79) | 23 (47.92) | 19 (70.37) | 67 (59.29) |
| Companions | 9 (23.68) | 20 (41.67) | 7 (25.93) | 36 (31.86) |
| Missing | 1 (2.63) | 2 (4.17) | 0 (—) | 3 (2.65) |
| **Respondent hosts their own blogs** | 27 (71.05) | 28 (58.33) | 15 (55.55) | 70 (61.94) |

## Motivations for blog use

Cluster 1 had slightly higher mean scores for using cancer blogs to seek cancer knowledge. Cluster 2 more frequently used blogs to express their own opinions. Across the 3 clusters in this study, encouraging others and sharing personal cancer stories were the primary motivators for blog use (mean score = 4.33 and 4.24, respectively). In cluster 1, seeking a second opinion, looking for timely updated information, and looking for compiled cancer information were the least motivating factors (mean scores = 2.84, 3.11, and 3.13, respectively). In addition, cluster 1 had the lowest mean score for expanding cancer knowledge (3.92) and cluster 2 had the lowest mean score for validating information (3.04).

## Behavioral changes

Cluster 1 members encountered fewer limitations for using cancer blogs than the other two clusters. Members in clusters 2 and 3 indicated that poor searching functions restricted their participation in cancer blogs, while members in cluster 1 indicated less agreement with that statement.

## Summary of cluster characteristics

Based on analysis of the data as determined by interpretation of survey results, Table 3 summarizes major characteristics of the 3 clusters found in this study. Cluster 1 (n = 38, 33.63%) was more likely to include new bloggers who were motivated to seek compiled information and were frequent online information

**Table 2**
Means of blog use, motivations, and behavioral changes

| | Mean | | | |
|---|---|---|---|---|
| | Cluster 1 (n = 38, 33.63%) | Cluster 2 (n = 48, 42.48%) | Cluster 3 (n = 27, 23.89%) | Total (n = 113, 100%) |
| **Uses** | | | | |
| Sought health care providers | 4.11 | 3.63 | 3.89 | 3.85 |
| Sought medical libraries or patient education centers | 2.66 | 2.92 | 2.56 | 2.74 |
| Sought family, relatives, friends | 3.11 | 2.96 | 3.11 | 3.04 |
| Sought mass media | 3.05 | 3.00 | 3.07 | 3.04 |
| Sought online discussion groups | 3.32 | 3.33 | 3.44 | 3.35 |
| Read blogs by others | 3.63 | 3.54 | 3.74 | 3.62 |
| Post comments on other blogs | 2.74 | 2.69 | 2.56 | 2.67 |
| Months of blog use | 16.76 | 23.46 | 21.19 | 20.66 |
| Self-evaluation of Internet skills | 4.44 | 4.35 | 4.07 | 4.31 |
| Total average* | 3.38 | 3.30 | 3.31 | 3.33 |
| **Motivations for primary use** | | | | |
| Seek timely updated information | 3.11 | 3.13 | 3.22 | 3.14 |
| Seek compiled cancer information | 3.13 | 2.98 | 2.78 | 2.98 |
| Communicate with others | 4.03 | 3.63 | 3.93 | 3.83 |
| Document personal cancer conditions | 4.05 | 4.00 | 4.07 | 4.04 |
| Share personal cancer stories | 4.18 | 4.17 | 4.44 | 4.24 |
| Seek encouragement | 4.11 | 3.96 | 4.11 | 4.04 |
| Encourage others | 4.26 | 4.25 | 4.56 | 4.33 |
| Read alternative information | 3.53 | 3.67 | 3.74 | 3.64 |
| Express own opinions | 4.03 | 4.06 | 4.04 | 4.04 |
| Receive emotional support | 4.05 | 3.83 | 4.15 | 3.98 |
| Expand cancer knowledge | 3.92 | 3.96 | 4.22 | 4.01 |
| Seek information to help friends and family | 3.71 | 3.50 | 3.56 | 3.58 |
| Seek second opinion | 2.84 | 2.90 | 2.78 | 2.85 |
| Validate information received from provider | 3.24 | 3.04 | 3.44 | 3.20 |
| Prepare background information for provider visit | 3.26 | 3.15 | 3.44 | 3.26 |
| Total average | 3.70 | 3.62 | 3.77 | 3.68 |
| **Behavioral changes** | | | | |
| Helped inform better cancer care decisions | 3.26 | 3.37 | 3.29 | 3.32 |
| Changed current self care options | 2.30 | 3.03 | 3.03 | 2.78 |
| Discussed information read from blog with physician | 2.81 | 3.39 | 3.48 | 3.22 |
| Prevented me from developing cancer | 2.08 | 2.43 | 2.88 | 2.42 |
| Decreased cost of health care | 2.09 | 2.50 | 2.81 | 2.44 |
| Decreased length of hospital stay | 2.26 | 2.55 | 3.04 | 2.57 |
| Found alternative methods | 2.51 | 3.00 | 3.35 | 2.92 |
| Coped with cancer situations | 4.32 | 4.23 | 4.15 | 4.24 |
| Were given a feeling of empowerment | 3.77 | 4.01 | 3.92 | 3.91 |
| Helped express frustration | 3.92 | 4.30 | 4.52 | 4.23 |
| Were satisfied with blog information | 3.82 | 3.73 | 3.78 | 3.77 |
| Learned new info | 3.98 | 4.06 | 4.07 | 4.04 |
| Formed relationships | 3.30 | 3.69 | 3.98 | 3.63 |
| Solidified existing relationships | 3.66 | 3.75 | 3.74 | 3.72 |
| Helped from further developing other diseases | 2.10 | 2.53 | 3.00 | 2.50 |
| Helped establish healthy lifestyle | 2.37 | 2.98 | 3.14 | 2.81 |
| Total average | 3.03 | 3.35 | 3.51 | 3.28 |
| **Limitations to blog use** | | | | |
| Lack of Internet connection | 1.53 | 2.15 | 3.12 | 2.17 |
| Lack of technical knowledge | 1.20 | 2.12 | 2.72 | 1.96 |
| High volume of messages posted | 1.58 | 2.40 | 2.96 | 2.26 |
| Language used is not understandable | 1.11 | 2.23 | 2.82 | 1.99 |
| Lack of credibility | 1.97 | 2.52 | 3.11 | 2.48 |
| Lack of security | 1.47 | 2.48 | 2.99 | 2.26 |
| Difficult navigation | 1.47 | 2.44 | 2.81 | 2.20 |
| No search function | 1.63 | 2.65 | 3.09 | 2.41 |
| Poor visual presentation | 1.56 | 2.48 | 3.18 | 2.34 |
| Total average | 1.50 | 2.39 | 2.98 | 2.23 |

* Does not include average months of blog use.

seekers. Cluster 2 (n = 48, 42.48%) was more likely to include long-time cancer blog users who also use traditional sources for information seeking. Individuals in cluster 3 (n = 27, 23.89%) were highly motivated and sought medically related information. In addition, bloggers in cluster 3 made the most frequent behavioral changes while using cancer blogs. Further details are described in Table 3.

## DISCUSSION/CONCLUSION

This study used an agglomerative, hierarchical clustering method to classify characteristics of unique groups among cancer bloggers. Employing this technique was most appropriate given that there was limited prior knowledge of the underlying structure and nonhierarchical methods could not clearly and objec-

**Table 3**
Summary of cluster characteristics*

| Cluster | Type of bloggers | Characteristics |
|---|---|---|
| **Cluster 1 (n = 38, 33.63%)** | New bloggers, motivated for compiled information, least influenced, frequent online information seekers | ■ Had the fewest months of blog use<br><br>■ Frequently commented on blogs by others<br>■ Had the highest self-efficacy reported<br>■ Sought health care providers for cancer information<br>■ Read blogs for compiled cancer information and to seek help for others<br>■ Read blogs to communicate with others<br>■ Were the least affected for changing care options<br>■ Had the least influence on preventing cancer<br>■ Had the smallest decrease in cost and length of hospital stay<br>■ Were highly satisfied with blog information<br>■ Rarely had limitations of blog use |
| **Cluster 2 (n = 48, 42.48%)** | Long-timers, traditional source seekers | ■ Had the most months of blog use<br>■ Sought medical libraries for cancer information<br>■ Read blogs to express own opinions and to seek second opinion<br>■ Were least interested in reading blogs to seek encouragement<br>■ Received strongest empowerment through blogs<br>■ Solidified existing relationships<br>■ Had mediocre level of limitations of blog use |
| **Cluster 3 (n = 27, 23.89%)** | Highly motivated group, medically related information seekers, most frequent behavioral change group | ■ Sought mass media and online discussion groups for cancer info<br>■ Had the most frequently read blogs by others and the least frequently posted comments on blogs by others<br>■ Read blogs to get timely updated information, to expand cancer knowledge, and to prepare background information<br>■ Read blogs to encourage others, to receive emotional support, and to share personal cancer stories<br>■ Obtained greatest benefits to change current care options, to further discussion with physicians, to prevent cancer, and to find alternative care<br>■ Found blogs helpful for expressing frustration and forming relationships<br>■ Had the highest score on the limitations of blog use |

* Characteristics grouped and assessed by author interpretation of mean scores in Table 2.

tively determine the number of clusters in a data set. Moreover, this method is useful when a study is still in its exploratory phase.

Some demographic data found in this study were different from previously reported data. Results from the analysis illustrated a dominant demographic group across all clusters: highly educated Caucasian females. This demographic group is inconsistent with other research findings about online cancer information seekers [5]. In this study, the bloggers were also older (average age fifty-seven) than in other studies. As older patients and their friends and family seek health-related issues (especially cancer), there are potential roles for using this new technological medium to deliver cancer information.

The study findings suggest that blogs are used more frequently to share emotional support and personal stories than medical knowledge, thus agreeing with reports and research that indicate blogs have gained their popularity over the past few years by supporting personal narratives, political commentaries, or accounts of personal experiences. This study confirmed the findings of previous research that suggested the use of blogs can lead cancer patients and their companions to engage in meaningful conversation and that sharing personal experiences via blogs may help

patients better cope with their cancer-related health conditions [6].

These results can inform the design of cancer blogs that provide customized (or personalized) assistance depending on the category of cancer blog users and their distinct characteristics. For instance, more attention by cancer information specialists (including medical librarians) might be given to people in cluster 1 because their motivation is to expand their knowledge about cancer as compared to people in cluster 2 who use blogs to seek emotional support.

Additionally, medical librarianship should not overlook bloggers and their uses, because blogs can be used as a health communication medium to disseminate cancer health information. In this sense, further in-depth analysis of cancer blog messages, including both posts and comments, may be beneficial in providing subject categorization to unorganized blog contexts. Medical librarians can, thus, play a key role in making information on blogs more easily assessable.

While this study provides useful findings, it has some limitations. The study sample is small and includes primarily Caucasian patients with bachelor's degrees. In addition, the use of convenience sampling and self-reported data could present bias in the reported results. Future studies should target a larger

pool of participants in a longitudinal setting for more valid, reliable, and generalizable findings.

## REFERENCES

1. Herring SC, Kouper I, Scheidt LA, Wright E. Women and children last: the discursive construction of Weblogs. [Web document]. [cited 28 Jun 2006]. <http://blog.lib.umn.edu/blogosphere/women_and_children.html>.
2. Hillan J. Physician use of patient-centered Weblogs and online journals. Clin Med Res 2003;1(4):333–5.
3. Mishne G. Experiments with mood classification in blog posts. In: Style2005: The 1st Workshop on Stylistic Analysis of Text for Information Access, SIGIR 2005; Salvador, Brazil; Aug 15–19, 2005.
4. Lenhart A, Horrigan J, Fallows D. Content creation online. [Web document]. Pew Internet & American Life Project, Feb 2004. [cited 20 Feb 2006]. <http://www.pewinternet.org/pdfs/PIP_Content_Creation_Report.pdf>.
5. Rainie L. The state of blogging. [Web document]. Pew Internet & American Life Project. [cited 28 Jun 2005]. <http://www.pewinternet.org/pdfs/PIP_blogging_data.pdf>.
6. Oransky I. Cancer blogs. Lancet Oncol 2005;6(11):838–9.
7. Harper C. Blogging and journalistic practice. Paper presented at: Association of Internet Researchers Annual Conference; Chicago, IL; 5–9 Oct 2005.
8. Anjewierden A, Hoog DE, Brussee R, Efimova L. Detecting knowledge flows in Weblogs. Presented at: 13th International Conference on Conceptual Structures. [Web document]. [cited 13 Jan 2006]. <http://anjo.blogs.com/metis/articles/kflows-pub.pdf>.
9. Cayzer S. Semantic blogging and decentralized knowledge management. Commun ACM 2005;47(12):47–52.
10. Gumbrecht M. Blogs as protected space. [Web document]. Presented at: WWW 2004: Workshop on the Weblogging Ecosytem: Aggregation, Analysis, and Dynamis. [cited 3 Mar 2006]. <http://www.blogpulse.com/papers/www2004gumbrecht.pdf>.
11. Kahn R, Kellner D. New media and Internet activism: from the battle of Seattle to blogging. New Media Soc 2004;6(1):87–95.
12. Google Blog Search. [Web document]. Google. [cited 2 May 2007]. <http://www.blogsearch.google.com>.
13. Johnson JD. Cancer-related information seeking. Cresskill, NJ: Hampton Press, 1997.
14. White MD. Questioning behavior on a consumer health electronic list. Libr Q 2000;70(3):302–34.
15. Papacharissi Z, Rubin AM. Predictors of Internet use. J Broadcast Electron Media 2000 Spring;44:175–96.
16. ClustanGraphics features. [Web document]. ClustanGraphics. [cited 23 Mar 2006]. <http://www.clustan.com/clustangraphics_features.html>.
17. Chen HM, Cooper MD. Using clustering techniques to detect usage patterns in a Web-based information system. J Am Soc Inf Sci Technol 2001;52(11):888–904.
18. Frank RE, Green PE. Numerical taxonomy in marketing analysis: a review article. J Mark Res 1968;5:83–98.
19. Milligan GW. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika 1980;45:325–42.
20. SPSS. SPSS base 13.0 user's guide. Upper Saddle River, NJ: Prentice Hall, 2004.
21. Mojena R. Hierarchical grouping methods and stopping rules: an evaluation. Comput J 1977;20:359–63.
22. Shaaban S, McKechnie J, Lockley S. Modeling information seeking behavior of AEC professionals on online technical information resources. ITcon 2003;8:265–81.

## AUTHORS' AFFILIATIONS

**Sujin Kim, PhD** (corresponding author), sujinkim @uky.edu, Assistant Professor, 518K, King Library South, School of Library and Information Science and Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY 40506-0039; **Deborah S. Chung, PhD,** dchung@uky.edu, Assistant Professor, 215 Grehan Building, School of Journalism and Telecommunications, University of Kentucky, Lexington, KY 40506-0042

# Expanding medical library support in response to the National Institutes of Health Public Access Policy

## Molly C. Barnett, MLS, AHIP; Molly W. Keener, MLIS

See end of article for authors' affiliations.

## INTRODUCTION

Responding to recent changes in the scholarly publishing process, Coy C. Carpenter Library is expanding its scholarly communications program to better support the research publication efforts of the faculty at Wake Forest University Health Sciences (WFUHS). Recent advances in open access publishing and archiving initiatives, adoption of the US National Institutes of Health (NIH) ''Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research'' (Public Access Policy) in 2005, the rapidly increasing pool of published biomedical research, rising costs of subscription rates, and continued barriers to access have necessitated an internal redesign of the library's Faculty Publications (FP) database. Changes in the scholarly publishing environment have also spurred the creation of online resource lists specifically addressing common issues in scholarly communications, including copyright and intellectual property ownership, open access, and the importance of scholarly publishing [1–3].

These efforts, coupled with plans for educational sessions on open access and copyright retention for faculty, are intended to address common questions raised during the publishing process. In particular, the FP database will bridge faculty publication citations to individuals' personnel profiles in the university's human resources department's management software, PeopleSoft, and to the full text of faculty-authored journal articles, thus providing the institution with a more complete picture of WFUHS faculty research initiatives and outcomes. This paper illustrates key objectives in Carpenter Library's strategy for supporting scholarly communications through enhancing the knowledge management applications of the FP data-

**Table 1**
2004–2006 comparison of totals by publication type

| Publication type | 2004–2005 | 2005–2006 | % change |
|---|---|---|---|
| Books | 10 | 9 | −10 |
| Chapters | 81 | 111 | +37 |
| Articles | 1,137 | 1,187 | +4 |
| Abstracts | 675 | 621 | −8 |
| Miscellaneous (editorials, letters) | 178 | 125 | −30 |
| Total | 2,081 | 2,053 | −1 |

base and leveraging the database's functions to promote open access to research.

## BACKGROUND

Since 1977, Carpenter Library has been responsible for collecting and organizing publication citations of WFUHS faculty. The goal of this effort has been to support the research and publishing activities of the faculty, as well as provide information to the Office of the Dean of the Wake Forest University School of Medicine, a division of WFUHS, for faculty promotions and advancement. Since the beginning, reports were created to track publication activities of faculty. The paper-based system was automated in 1988 as the FP database in Cuadra Star, and in 1991 the information was moved to MS Access. Only published, scholarly materials—articles, abstracts, book reviews, chapters, books, editorials, and letters—are included in the database. Until recently, citations from the database were formatted and printed as an appendix of the Dean's Annual Report (DAR). Now the library maintains a searchable Web interface for departments as well as visitors. Statistics are compiled for the DAR by document type totals (Table 1) and by department.

## PROCESS

As mandated by the office of the dean, faculty must submit all publication information, including the bibliographic citation and a copy of the first page of the material, to the library. The library maintains a Faculty Publications Web page with submittal forms, guidelines for submission, a link to the previous year's DAR, and library contact information [4]. Approximately 60% of the publications are received directly from faculty, with the remaining 40% gathered by library staff from bibliographic databases such as PubMed, BIOSIS, and Web of Science. The staff verifies all submitted materials and has maintained very high standards of accuracy over the years. Bibliographic information is entered into the database, and a Web-based interface is available for departments to track submittals and as a research tool for users.

In 2005, a committee, including library representation, was formed to create a Faculty Information (FI) database for the school of medicine. The database would include biographical and professional information for faculty. The goal of the project is to connect faculty profiles with research interests, grants, internal protocols, and publications, which would be easily accessible to authorized users. WFUHS has used PeopleSoft software for vital information, salary data, and personnel activity for many years, and the committee decided to integrate professional information such as research interests, grants, and publications into a unified source. To this end, a team composed of library and information technology staff created a new interface for faculty publication data entry utilizing PeopleSoft software, which links information in the FI database through employee identification numbers. Figure 1 shows the FP database input screen, still under development, which includes fields linking users to full-text articles via the digital object identifier (DOI), uniform resource locator (URL), or the unique identifier used in PubMed Central (PMCID).

## NEW DEVELOPMENTS

In light of rapid increases in both the volume of scholarly research articles published annually and the subscription costs of scientific, technical, and medical journals [2, 3, 5], momentum is building in the scholarly communications community for broader, less restrictive access to vital scientific and biomedical research information, particularly to taxpayer-funded research [5–8]. NIH, at the behest of the US Congress, adopted the Public Access Policy on May 2, 2005 [9]. This policy requests that copies of all peer-reviewed scholarly articles resulting from NIH-funded research projects be archived in PubMed Central (PMC) [10]. Created in February 2000, PMC is a freely accessible, full-text digital repository of peer-reviewed scholarly articles from biomedical and life sciences journals [9]. In September 2006, approximately 700,000 articles were in PMC [11]. This figure includes both materials submitted as a result of the NIH policy and articles submitted by publishers.

Although some WFUHS faculty members have complied with the NIH Public Access Policy's archiving request, the library advocates increasing the percentage of archived publications. The institutional contribution rate (ICR), modeled on a search query developed at the Health Sciences Library and Informatics Center, University of New Mexico [12], for WFUHS is 4.31%.* This figure takes into account archived author manuscripts in PMC published on or after the May 2, 2005, NIH Public Access Policy adoption date through May 7, 2007 (n = 20 for WFUHS) and the number of eligible articles cited in PubMed published during the same time period (n = 464 WFUHS-authored articles resulting from NIH-funded research). This compliance rate is disproportionate to the number of publications likely to result from the 1,279 NIH awards granted to WFHUS faculty during the past 5 fiscal years (July 1,

---

* Affiliation can alternatively be stated as Wake Forest University Health Sciences, Wake Forest University School of Medicine, or Wake Forest University Baptist Medical Center, and the WFUHS campus has a separate zip code from the main campus, so the search query started: ((("wake forest" AND (health OR medicine OR medical)) OR 27157) NOT (27106 OR 27109)).

**Figure 1**
Faculty Publications database input screen (under development)



2001–June 30, 2006), which account for 68.29% of all external funding received by WFUHS researchers during that time [13, 14].†

## FUTURE OBJECTIVES

The inability to accurately track the number of publications stemming from individual grants should im-

prove with the changes to and increased functionality of the FP database. Currently, articles in the FP database include a DOI link to articles regardless of whether the full text is available via a WFUHS library subscription. The FP staff plans to include the PMCID in each relevant record to allow users to link to the freely accessible full text of faculty research articles and, in turn, encourage submission. Additionally, users will be able to link from the citation to other information in the FI database such as grant information, research protocols, and funding data. This information is useful, if not required, when applying for and renewing federal grants, as well as filing grant progress reports. The database should also fill a valuable role in promoting the exchange of internal knowledge about research interests and findings and potentially enabling connections among researchers.

† Data exclude subcontract awards. Due to the length of time typically needed to produce publishable research results and allowing for varying time periods needed to complete the prepublication peer-review process, data for the past five fiscal years is given with the reasonable assumption that most National Institutes of Health (NIH)–funded research projects supported during that time period would be eligible for author manuscript archiving in compliance with the NIH Public Access Policy.

## PROMOTING OPEN ACCESS

The library recognizes that current low rates of compliance with PMC archiving might be due to confusion about and general lack of awareness of the NIH Public Access Policy and is tailoring its scholarly communications program to better support faculty research endeavors. Previously created ''toolkits,'' lists of online and print resources appearing on the library's Website, have been updated and are now centrally located in the site's Scholarly Publishing Assistance section. Resource pages relevant to scholarly communications include the NIH Public Access Policy, scholarly publishing and open access, copyright and intellectual property, and scientific writing [15]. The need for straightforward resource pages was identified in an interdepartmental meeting hosted by the library with representatives from the office of the dean and the office of research.

Beyond the redesigns of the FP database and the Scholarly Publishing Assistance section of the library's Website, direct educational outreach programs are under development. Plans include ''lunch and learn'' seminars to be delivered to faculty in individual departments, during which time a librarian will give a ten-to-fifteen-minute presentation on topics including open access archiving, copyright permission and retention, the NIH Public Access Policy and PMC, current compliance rates, and ways these issues impact researchers. Time for questions will be allowed following the presentation, with the total estimated time of the program to be thirty minutes. By taking these seminars to the various departments directly and limiting the time of the program to half an hour, library staff hope that many faculty will be willing to attend.

## CONCLUSION

Continued growth of the scholarly communications program at Carpenter Library will be dictated by the success of current and future initiatives undertaken by the library, as well as by changes in the broader scholarly communications community. Potential adoption of open access archiving requirements stipulated by either funding sources or government bodies will also guide library staff in prioritizing their efforts to support faculty researchers. Enhancement of the FP database and the anticipated interconnectivity with the FI database should facilitate increased functionality and accountability in faculty's and administrators' research tracking efforts. Libraries interested in highlighting the research achievements of faculty but lacking the resources necessary to build and maintain an institutional repository might consider creating a bibliographic database of faculty publication citations with links to full text available elsewhere. Librarians will be able to use these databases to engage faculty and administrators in dialogue about the advantages of open access and compliance with the NIH Public Access Policy and the necessity of understanding copyright, as well as provide quantifiable evidence of the knowledge congregated in their parent institutions. These efforts further underscore libraries' valuable roles in facilitating the expansion of scientific information.

## REFERENCES

1. Guterman L. The promise and peril of ''open access.'' Chron High Educ 2004 Jan 30;50(21):A10–4.
2. Watson LA, Login IS, Burns JM. Exploring new ways of publishing: a library-faculty partnership. J Med Libr Assoc 2003 Apr;91(2):245–7.
3. Koehler BM, Roderer NK. Scholarly communications program: force for change. Biomed Digit Libr [serial online]. 2006 Jun 21;3(6). [cited 4 May 2007]. <http://www.bio-diglib.com/content/3/1/6>.
4. Coy C. Carpenter Library. Faculty publications. [Web document]. Winston-Salem, NC: The Library, 2007. [rev. 2 Feb 2007; cited 24 Apr 2007]. <http://www1.wfubmc.edu/Library/About+the+Library/Library+Departments/Faculty+Publications/>.
5. Yiotis K. The Open Access Initiative: a new paradigm for scholarly communications. Inf Tech Libr 2005 Dec;24(4):157–62.
6. Bergman SS. The scholarly communication movement: highlights and recent developments. Collect Build 2006; 25(4):108–28.
7. Johnson RK. Will research sharing keep pace with the Internet? J Neurosci 2006 Sep 16;26(37):9349–51.
8. Van Orsdel LC, Born K. Serial wars: as open access gains ground, STM publishers change tactics, and librarians ask hard questions. Libr J 2007 Apr 15;132(7):43–8.
9. National Institutes of Health. NIH public access: questions & answers. [Web document]. Bethesda, MD: The Institutes, 2007. [cited 19 Apr 2007]. <http://publicaccess.nih.gov/publicaccess_QandA.htm>.
10. National Institutes of Health. PMC overview. [Web document]. Bethesda, MD: The Institutes, 2007. [rev. 7 Jan 2005; cited 4 May 2007]. <http://www.pubmedcentral.nih.gov/about/intro.html>.
11. National Library of Medicine. Minutes of the Board of Regents: September 19–20, 2006. [Web document]. Bethesda, MD: The Library, 2006. [cited 3 May 2007]. <http://www.nlm.nih.gov/od/bor/9-06bor.pdf>.
12. Kroth PJ, Aspinall EE, Phillips HE. The National Institutes of Health (NIH) Policy on Enhancing Public Access: tracking institutional contribution rates. J Med Libr Assoc 2006 Jul;94(3):279–83.
13. National Institutes of Health, Office of Extramural Research. Award trends. [Web document]. Bethesda, MD: The Institutes, 2007. [cited 15 Mar 2007]. <http://grants.nih.gov/grants/award/awardtr.htm>.
14. Office of Research, Wake Forest University Health Sciences. FY06 sponsored research by funding source and annual comparison. [Web document]. Winston-Salem, NC: The University, 2006. [cited 15 Mar 2007]. <http://www1.wfubmc.edu/NR/rdonlyres/250FC563-535B-4925-9659-56EA577F5445/0/FY06TABLE12.DOC>.
15. Coy C. Carpenter Library. Scholarly publishing assistance. [Web document]. Winston-Salem, NC: The Library, 2007. [rev. 5 Apr 2007; cited 24 Apr 2007]. <http://www1.wfubmc.edu/Library/Research+Publishing+Resources/index.htm>.

## AUTHORS' AFFILIATIONS

**Molly C. Barnett, MLS, AHIP** (contact author), mbarnett@wfubmc.edu, Associate Director of Technical Services; **Molly W. Keener, MLIS,** mkeener@wfubmc.edu, Reference Librarian; Coy C. Carpenter Library, Wake Forest University Health Sciences, Medical Center Boulevard, Winston-Salem, NC 27157-1069