How Chemists are Really Finding and Using Information in our Digital Environment

Panelists:
Cecelia Brown, PhD
SLIS
The University of Oklahoma
401 W. Brooks, Room 120
Norman, OK 73019
Phone: (405) 325 3921
FAX: (405) 325 7648
e-mail: cbrown@ou.edu

Catherine Blake, PhD
SILS
UNC at Chapel Hill
214A Manning Hall CB#3360
Chapel Hill, NC 27599-3360
Phone: (919) 843 5636
FAX: (919) 962 8071
e-mail: cablake@email.unc.edu

Ashley Brown, MSLS
SILS
UNC Chapel Hill
214A Manning Hall CB#3360
Chapel Hill, NC 27599-3360
Phone: (919) 843 5636
FAX: (919) 962 8071
e-mail: Ashley_Brown@unc.edu

Moderator:
Carol Tenopir, PhD
School of Information Sciences
College of Communication and Information
University of Tennessee
1345 Circle Park Drive, Room 451
Knoxville, TN 37996-0341
Phone: (865) 974 7911
e-mail: ctenopir@utk.edu

Sponsors: SIG-STI and SIG-USE

Abstract
The quantity of information available to scientists continues to increase alarmingly with the top 418 journals more than 110,000 articles in 2002. Chemists are particularly prone to this information overload because they have more personal journal subscriptions (Noble & Coughlin 1997), read more articles (Tenopir et al. 2003), and access more journals than other scientists (Davis & Solla 2003). This quantity of information makes it difficult to stay current as well as explore new connections between articles. To ensure the provision and improvement of access to this essential element of scientific productivity and innovation it is critical that information scientists understand how the

print and digital information explosion is influencing the information behavior of scientists.

To elucidate the information sources chemists are using, two studies employing citation analysis of American Chemical Society (ACS) journals will be discussed. The third study employs interviews to describe the role of the literature in chemists' discovery processes. The goal is to provide a realistic snapshot of the current information behavior of chemists in order to assist in the creation, design, and delivery of information products and services for the support of the research, teaching, and creative activities of chemists and scientists in general.

Does Electronic Availability Enhance Use of Chemistry Information?
Panelist: Cecelia Brown

The advancement of science is dependent upon a complex flow of information extending from individual research laboratories to the global scientific community. As first noted over 30 years ago by R.K. Merton (1968), the primary mode of scientific communication is publication in peer-reviewed journals. This gold standard still prevails today, yet studies of the information behavior of scientists conducted during the past decade by Tenopir (2000), Hurd (1996), McCain (2000), and Brown (1999) have found scientists to also use a variety of digital information products and services. Brooks (2002) and Brown (2001) have found physicists to fully embrace e-prints as a viable and integral component of their scholarly discourse. In contrast, chemists are reluctant to accept the e-print as a viable communication mode due to the lack of peer-review of and confidence in the permanence of digital archives. In contrast, life scientists were found in studies by to be willing to freely share research related data in the myriad of publicly available web-based genomic and proteomic databases (McCain 2000; Brown 2003a). The biochemists and microbiologists interviewed are willing to share their data in public electronic forums for the "public good" while dissemination via peer-reviewed journals is seen as necessary for recognition in their disciplines. This paper explores how authors and readers of chemistry articles are employing electronic scientific information in their scholarly work using the unobtrusive measures of citation and content analyses of what chemists leave behind in the course of communicating their research (Case 2002).

The journal population selected for content and citation analyses was eight ACS titles from a range of fields of chemistry published during 1996, 2000, and 2004. Content analysis of the journals was conducted for the occurrence of websites both within the full text of the articles as well as the article reference lists. To understand how chemists use electronic information in their articles, the nature of the website content and the placement of the websites within the articles was noted. Additionally, as chemists voice concern over the stability of web information, the current viability of the websites used was assessed. Citation and content analyses capture article use in a very concrete and direct way, yet do not account for use in less obvious, and perhaps unrecorded, fashions such as mentally gleaning and storing information for later use. For the past two to three years the ACS website has published a listing of the 20 top accessed titles for each journal thereby providing an index of less direct article use. To discover if use of an article as indicated by citation counts relates to the number of times it is accessed, this study reports the results of a comparison of these two sets of data. Finally, to learn whether the purpose or content of an article dictates its levels of citation and/or access, the content of the titles of the top accessed and cited articles was also conducted.

Questions posed by this panelist are:
1. Are chemists incorporating freely available electronic information in their scholarly works?

2. What is the nature and stability of the freely available electronic information used?
3. Are chemists citing chemistry articles that incorporate electronic information?
4. Does the number of accesses of a chemistry article relate to its citation rate?
5. Does purpose or content of an article relate to citation and/or access levels?

Information Use Surrounding the Discovery Processes of Chemists and Chemical Engineers
Panelist: Catherine Blake

The quantity of information available electronically to scientists continues to increase at an alarming rate. For example, the top 418 journals in ISI (http://isiknowledge.com) published more than 110,000 articles in 2002 alone. Chemists are particularly prone to information overload because they have more personal journal subscriptions (Noble & Coughlin, 1997), read more articles than other scientists (Tenopir et al. 2003), and download and view far more journals than other subpopulations in the sciences (Davis & Solla 2003). Although published articles capture the latest findings in chemistry, the quantity of information makes it difficult for scientists to stay current, and they have little time to explore new connections between articles.

Text mining technology provides scientists with a new way to explore information reported in the published literature. Swanson's pioneering work in the 1980s resulted in the ABC model (Swanson & Smalheiser 1997), which identifies logical connections between bibliographically disjoint collections of medical articles. Using the ABC model, Swanson hypothesized connections between and magnesium and migraines (Swanson 1988); and fish oil and Raynaud's disease (Swanson 1986). The scientific literature at the time reported neither of these connections, yet both were later verified using clinical trials. The METIS text mining system extracts information from full-text articles, and provides the user with a visual representation of contradictions and redundancies (Blake 2005). Despite these and other examples in medicine, few studies have explored the application of text mining in chemistry literature. Further, we have yet to understand the extent to which text mining technology reflects how scientists arrive at new hypotheses.

We are in the second year of an NSF funded study to explore text mining in chemistry literature. The grant comprises both a user study component (the focus of this panel discussion) and the development of new text mining technology. The user study comprised face-to-face interviews with experienced scientists in chemistry and chemical engineering and extends our previous understanding (Blake & Pratt In Press[a&b]), of scientists' information behaviors as they write journal articles. We used a critical incident technique and asked the scientists to define discovery, describe how they arrive at a research question, and their current personal information management strategies. We also asked the scientists to outline the processes they use to write a manuscript using a set of ten cards that included discussing, experimenting, and searching. The results of the user study provides key criterion that we are using in our design of our text mining system to facilitate discovery from chemistry literature.

Questions posed by this panelist are:
1. What processes do chemists and chemical engineers (CCE) use as they make new discoveries?
2. How do CCE currently incorporate electronic information into the discovery processes?
3. In what ways do existing retrieval systems support the information delivery during the discovery process and where are the gaps?
4. What key criterion should we take from this user study as we design and develop text mining systems that facilitate discovery from chemical literature?

Using Citation Analysis to Inform Collection Development
Panelist: Ashley Brown

Citation Analysis, in its most traditional sense is the identification and tabulation of the number of times a work (article, monograph, etc.) is utilized as a reference by other authors. Intertwined with this idea of citation analysis is the use of the *Institute of Scientific Information's* (ISI) journal impact factor. This information is often utilized by academic libraries and research institutions for selection and deselection of serials titles as well as evaluation for faculty tenure.

This panelist will discuss a citation analysis using articles from 27 journals published by the American Chemical Society between 2000 through 2004. Our analysis compares and contrasts citation usage between Chemistry faculty at Duke University, North Carolina State University and the University of North Carolina at Chapel Hill. We also compare the impact of journal self-citation, such as when article x is published in journal y and cites journal y with the ISI's impact factor within the corpus of available titles.

In addition to campus comparisons, we compare our results with the existing ISI journal impact factors and, when available, citation and publication analysis purchased by each institution. In addition to the alterations in the impact formula, we have included the publication year of the cited article. This date is particularly important during collection development because it can reveal differences over time, and suggest titles for purchase or cancellation to understand whether a citation refers to an article published in 1949, for example, or from the previous year. This additional information allows the selector to determine whether a purchase of current material or archival material supports the institutional research more effectively.

We conducted this study to explore how the ISI impact factor changes based on different institutions and over time. Although we do not have access to the same collections of articles at ISI, we are confident the more than 100,000 full text articles and more than 3.2 million references will provide an interesting comparison for both practitioners and researchers. Our analysis will explore the variability with respect to self-publication, and over time. The time aspect, which is not included in the existing ISI subscriptions can provide valuable information for collection development. For example if a journal is no longer frequently cited the reference collection should focus on back-issues and not on current subscriptions.

Questions posed by this panelist are:
1. How can citation information be used for collection development?
2. How stable are citation patterns over time and for different user populations?
3. How can citation data provide a more detailed analysis of collection use?
4. What can the practitioner learn through citation analysis?
5. What is the continued value of ISI data?

References
Blake, C. (2005). Information Synthesis: A New Approach to Explore Secondary Information in Scientific Literature*. Proceedings of the 5th ACM/IEEE-CS Joint conference on Digital libraries*, Denver, CO. 434: 56-64.

Blake, C. & Pratt, W. (In Press[a]). Collaborative Information Synthesis I: A Model of Information Behaviors of Scientists in Medicine and Public Health. *Journal of the American Information Society of Science and Technology*.

Blake, C. & Pratt, W. (In Press[b]). Collaborative Information Synthesis II: Recommendations for Information Systems to Support Synthesis Activities. *Journal of the American Information Society of Science and Technology*.

Brooks, T. (2002). Peripatetic preprints in particle physics. *Proceedings of the 65th Annual Meeting of the ASIST*, Philadelphia, PA. 39, 452–453.

Brown, C.M. (1999). Information seeking behavior of scientists in the electronic information age: Astronomers, chemists, mathematicians, and physicists. *Journal of the American Society of Information Science,* 50(10): 929-943.

Brown, C.M. (2001). The e-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society of Information Science and Technology,* 52(3):187-200.

Brown, C.M. (2003a). The changing face of scientific discourse: analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society of Information Science and Technology,* 54(10): 926-938.

Case, D.O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior.* San Diego, CA: Academic Press p. 207.

Davis, M. P., & Solla, R. L. (2003). An IP-Level analysis of usage statistics for electronic journals in chemistry: Making inferences about user behavior. *Journal of the American Society of Information Science and Technology, 54*(11), 1062-1068.

Hurd, J. M., Wheeler, A. C., & Curtis, K. L. (1996), Information seeking behavior of faculty: Use of indexes and abstracts by scientists and engineers. *ASIS Proceedings,* 29: 136-143.

Merton, R. K. (1968). The Matthew Effect in science. *Science*, 159: 56-63.

McCain, K. W. (2000). Sharing digitized research-related information on the World Wide Web. *Journal of the American Society of Information Science and Technology,* 51(14): 1321-1327.

Noble, L. R., & Coughlin, C. (1997). Information-seeking practices of Canadian academic chemists: A Study of Information Needs and Use of Resources in Chemistry. *Canadian Journal of Communication, 22*(3/4): 1-10.

Smith, E.T. (2003). Changes in faculty reading behaviors: The impact of electronic journals on the University of Georgia. *Journal of Academic Librarianship,* 29(3): 162-168.

Swanson, D. R. (1986). Fish-oil, Raynaud's Syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine,* 30(1): 7-18.

Swanson, D. R. (1988). Migraine and magnesium: eleven neglected connections. *Perspectives in Biology and Medicine,* 31: 526-557.

Swanson, D. R., & Smalheiser, N. R. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence,* 91: 183-203.

Tenopir, C., & D.K. King. (2000). *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers* (p. 91). Washington, DC: SLA Publishing.

Tenopir, C., King, D. W., Boyce, P., Grayson, M., Zhang, Y., & Ebuen, M. (2003). Patterns of journal use by scientists through three evolutionary phases. *D-Lib Magazine, 9:* 1-15.