

Academic Research in the Cyberspace Age

By Susan Willey

For the academician, accessing knowledge is of crucial importance and library databases are increasingly important as research tools. But I fear we in the professoriate may be naive about what information we are being allowed to access on the academic databases we rely on for our research.

I've made a number of disturbing discoveries in my own research about information limitations, access, and market controls that have serious implications for all faculty research. Yet rarely have I come across commentary on what I see as the *real* issue of our cyberspace age—the marketing of knowledge. Before I relate my research journey, let me provide a brief background about electronic information systems.

Massive amounts of information are stored on what *Online Magazine* calls “mega-databases.”¹ These mega-databases include all the usual ones that we use in our

research—Info Trac, ProQuest, Wilson's OmniFile, Dialog, and Lexis-Nexis, for example. Most of these are aimed at the academic market and libraries across the nation pay hefty sums for them.²

But some people are wondering whether this information glut is backfiring. In his article, “Big Databases Pose Big Questions,” Mick O'Leary argues that: “We may be at the point of diminishing returns with these mega-databases. They stretch the envelope so far that we may be better off contracting it instead of expanding it further.”³

The problem is simple. These databases have cataloged so much information that people have difficulties accessing it. Despite the high sums libraries pay, students and faculty aren't using the information. Because each database's search methods differ, even the academically astute often are challenged to find what they are seeking.

O'Leary notes that databases are adding more and more informa-

Susan Willey is an assistant professor of journalism at Florida Atlantic University in Jupiter, Florida. She earned her doctorate from the University of Missouri School of Journalism in 1999 and her master's degree from the University of South Florida-St. Petersburg, where she was a Poynter Fellow. She was a research fellow at the Charles F. Kettering Foundation in Dayton, Ohio prior to assuming the position as journalism professor at FAU. She was an investigative and government reporter for more than 15 years.

One article that examined Web-based research paints a picture of concern about the use of search engines.

tion of “marginal value.” My experience suggests there may be an even bigger problem: Marketing restrictions are limiting researchers’ access to the information they may need. It is a paradox. We have much information that we don’t need and are denied the information we do.

Consider that the Internet has several different kinds of search engines to scan the approximate 3 million servers and 800 million indexable Web pages.⁴ In 1979, the first online database directory, profiled 300 publicly available databases. By 1992, the number grew to 5,300.⁵ The 2001 Gale Directory of Databases profiles 13,854 publicly available electronic databases.⁶

At first glance, the availability of massive amounts of information may seem wonderful for the academic researcher. Databases provide almost instant access to even the most obscure journals and articles that can complement our search for knowledge.

But, as Marshall McLuhan said decades ago, these new mediums are changing us and, I believe, changing our research as well. One of my former professors cautioned his students to remember that all research is a compromise. But the question I struggle with now is how to distinguish between research as compromise and compromised research.

Although there have been numerous studies and journal articles on the new media and the Internet, my concern is more mundane and, as a new faculty member, practical. I’m worried about actually *doing* the research and how to determine the validity and reliability of what I’m finding—or not finding—on the Web and on academic databases.⁷

Despite the abundance of articles on new media, there are surprisingly few studies that probe the idea of research itself. One article that did critically examine Web-based research paints a picture of concern about the use of search engines.

Steve Lawrence and C. Lee Giles found in their studies that about 85 percent of Web users use search engines to locate information. The search engine, obviously, is critically important in accessing information on the Web. These researchers found that different search engines retrieve different documents even when the same key words were used in the query.

In addition, the authors discovered that search engines index only a small fraction of the Web, most likely because it is not economically feasible to do more. Lawrence and Lee also found that the search engine databases were not always up to date.⁸

Another study found that most

The marketplace is dominated by a few mega-players who have immense control over the world's information.

Web users—including students and professors—don't understand how search engines work. The study discovered that "less than 5 percent of the users surveyed knew how to conduct effective web searches."⁹

A recent report from the National Education Association (NEA) revealed that two-thirds of the nation's faculty members find it very difficult to keep up with technology and efforts to do so result in considerable stress.¹⁰

Yet we continue to be intrigued by the possibilities of the new media technology and all the electronic databases. Those of us who once spent hours, weeks, or months trying to find articles and reports now have the Library of Congress at our keyboard fingertips. We are as close to needed information as our university and home computers. Through these mega-databases we can, with a few key words, find just about everything we want.

Or can we?

Don't count on it. Those who control information have their own priority—profit. The story frame for this endeavor is not "the pursuit of knowledge." It is, as one author wrote, "... truly a business story; it's about money and sales, real estate and corporate power."¹¹

Communication scholar Harold Lasswell's standard transmission

model of communication—Who says what, in what channel, to whom, with what effect?—may be worth examining in the context of the new "channels" of the Web and software databases.

For the researcher's purpose, the critical element in Lasswell's model may be the "Who." Just *who* is controlling the information we search? This confronts a basic question of research—"How do we know what we know?" But in this day of Internet access, various search engines, and different contracts for different databases, and mega-databases, this question—and the answers—become elusive.

One writer compares the "Database Marketplace 2001" with car racing. As it gathers more money and fans, it is moving out of control, expanding faster and faster. As these major database corporations merge, the information marketplace is increasingly dominated by a few mega-players that have immense control over the world's information.

As the author wrote: "The full impact of this dominance by a few hasn't set in.... Academic librarians in particular worry about the long-term impact of diminishing competition."¹² As a scholar, so do I.

What it comes down to is information control and money. My recent experience in data gathering has left me skeptical about

For the next five months, I became engrossed in doing research on why I couldn't do research.

research databases and concerned about faculty ignorance of the information controls that may have profound effects on current and future research.

Two years prior to receiving my doctorate from the University of Missouri School of Journalism, I was a pre-doctoral fellow at the Charles F. Kettering Foundation in Dayton, Ohio. One of my research interests is in journalism's relationship to democracy, and I focus on print journalism, primarily newspapers.

Because the Kettering Foundation is a research foundation, I had access to several databases in my dissertation research. Lexis-Nexis was one I used frequently because the "Nexis" part of Lexis-Nexis offers full-text access to numerous newspapers and periodicals. I was able to use this database to do some in-depth searches, narrowing the field to specific publications, calling up only page 1A stories on a certain date, limiting the articles to specific key words.

I took this kind of search ability for granted until I was offered a position as a faculty member at Florida Atlantic University in Boca Raton, FL. I was concerned that I would no longer have access to this incredible database, but university officials assured me that their libraries had access to the Lexis-Nexis database.

Once I settled into my office on FAU's newest campus in Jupiter, I began work on my follow-up research. But when I accessed the web-based Lexis-Nexis Academic Universe and attempted to use the same search pattern I had used with the Lexis-Nexis software and Web version at the Kettering Foundation, my search brought up zero hits.¹³ Knowing this could not be accurate, I sought help from the university's librarian. She, too, could not get any hits. There were no stories.

The access to specific articles was vital to my future research. Why could I access them on the foundation's Lexis-Nexis database, but not on the university's database? For the next five months, instead of doing the research I had intended to do, I became engrossed in doing research on why I couldn't do research.

What I discovered, I believe, has incredible implications for all research being done at academic institutions and spawns more questions about information controls, accessibility, search mechanisms, methodologies, and the trustworthiness of our research findings in this cyberspace age.

I began calling people and asking questions. Among those I interviewed were librarians from my

It was apparent from the start that the academic version of Lexis-Nexis fell far short of the corporate version.

alma maters, in Missouri and at the University of South Florida, where I had earned my master's degree, a law librarian at the University of Florida, and a friend who is a library researcher at the Poynter Institute for Media Studies in St. Petersburg. In addition, I called friends at the Kettering Foundation to make sure they could, indeed, access the information that I was unable to access. They could. Eventually, I was able to contact the Lexis-Nexis representative for my area, who confirmed my findings from the interviews.

What I found was that, unless people were in the position I was in—working for a research foundation before they went into the academy—they were unlikely to realize they were receiving different versions of Lexis-Nexis at a university library than that available at other institutions.

There are distinctly different databases, operating under similar names. The Kettering Foundation had an older software version of Lexis-Nexis, but now uses the Web-based version called “Lexis-Nexis Universe.” This is a corporate account that has more full-text databases and better search capabilities. The academic version—Lexis-Nexis *Academic* Universe—has much more limited search capabilities and fewer newspapers on the database.

A few of the university librarians I spoke with were unaware of this difference. In fact, some misused the name, calling their database Lexis-Nexis Universe, leaving out the “Academic” qualifier.

The Universe group of databases was developed by Lexis-Nexis for two distinct markets. The “Academic” Universe—sold to universities—is a “condensed” version of the main Lexis-Nexis service—“Universe.”¹⁴ It was apparent from the start that there was a vast difference and the academic version fell far short of the corporate version I was accustomed to using, and on which all my previous research was based. It was a revelation about how information can be, and is, manipulated by numerous control factors, ownership and money concerns. As one observer wrote: “Information is clearly the currency of the new economy.”¹⁵

When I spoke with the Lexis-Nexis representative, she was familiar with the Kettering Foundation because Lexis-Nexis is headquartered just down the road from the foundation offices in Dayton. She was blunt and honest, telling me that I could not expect to get the same search capabilities at a university as I had at the foundation.¹⁶ The corporate account is more expensive than an academic one,

I found it rather incredible that corporations had more and better access to databases than did universities.

has more newspaper databases and is capable of a much more detailed search that can manipulated in various ways to produce the kind of information I sought.

I was told by her, and by the librarians I spoke with, that Lexis-Nexis Academic Universe was never intended to do the kind of research that faculty members might do. It was designed primarily for undergraduates.¹⁷ The concept of utilitarianism seems alive and well in cyberspace and the electronic database world. Databases are created to provide the greatest good for the greatest number, in this case undergraduates.

I found it rather incredible that corporations had more and better access to databases than did universities and other institutions of higher learning where, supposedly, the research would be more likely to occur. Then I realized it was not a matter of logic but of money and information control.

The issue here was not research, but sales, and it included the newspapers themselves. Private, for-profit corporations and generously endowed, non-profit foundations often have more dollars to spend than publicly funded state academic institutions. Since the corporate versions of the databases are more expensive, it seems reasonable that the database service company can pay the newspa-

pers and periodicals more money under this contract.

There are other restrictions on faculty researchers, as well. Because of concerns about intellectual property rights, there's no way that faculty can legally access this expansive database. The Lexis-Nexis representative told me that, because of licensing restrictions, anyone connected with the academy is prohibited from purchasing the corporate version of Lexis-Nexis. This means that, even if university faculty wanted to spend the money to obtain access to better databases and search engines, they would not be allowed to do so.

This is a severe handicap for faculty researchers. In terms of newspaper research, there are many and distinct differences between the academic and corporate versions of Lexis-Nexis. In addition to the lack of access, or certain newspapers providing only selected articles, the two versions perform searches differently.

For example, under the corporate Lexis-Nexis Universe, the search engine automatically includes the newspaper section name in the search. This means that, if I wanted to search the "Religion Section" of any newspaper, all I would have to do is type in the keyword "religion" for that specific newspaper's database and all the

How can one describe the methodology when different search engines won't perform the same way?

stories in that section would be called up.

In the academic version, the section heads are not included in the searchable areas so when the word, "religion" is keyed in, the search engine only selects those articles that have the actual word, "religion," in the body of the text itself. Hence I received zero hits. Can you begin to see the problems here?

It gets worse. Some newspapers have rejected offers from Lexis-Nexis to allow their copyrighted material to be placed on the database or, in some cases, they've withdrawn their material only from the academic version. This means that one newspaper might provide current, full-text articles on the corporate Lexis-Nexis Universe, but it may not provide any on Lexis-Nexis Academic Universe.

This was my difficulty. The newspaper I was researching no longer offered its publication to the academic version of Lexis-Nexis, which was why I was receiving zero hits. Some of these research problems may be out of Lexis-Nexis' control. If a newspaper decides it does not want to sell its articles to the database, but instead decides to create its own search engines to provide ways to purchase the articles on their own newspaper Web site, it certainly has that freedom.

In some cases, certain newspa-

pers pre-select articles they will sell to the database. These newspapers then are noted as having a "selected" database. This means that a researcher has no idea what else may have been published that she might need. Instead of the researcher selecting the copy to study, she only sees a pre-selected grouping that she can examine. The population has already been chosen—but not by the researcher.

For a journalist turned academician, this is frightening. What am I not seeing? What am I not being allowed to access? Whose voices are not being heard because someone somewhere decided not to include them?

And it isn't only the philosophical questions that are of concern. It's a matter of practicality. Any scholar knows that she must describe the methodology used in her research so that, if another scholar chooses to replicate it, she can.

How can one describe the methodology when different search engines won't perform the same way? For example, you can do research using certain key words and connectors and date limitations on the Lexis-Nexis database that do not correspond to the search engines of private newspaper Web pages. In other words, the research method must be changed.

If you cannot compare across search engines, then how valid and reliable is the research?

In the final analysis, the question is whether the documents obtained through different methodologies and search strategies are comparable to other searches? If you can't compare across search engines, then how reliable is the research? One solution, I suppose, is to relinquish your control as the researcher and let the newspaper's search engine determine what methodology you can use.

Lawrence and Giles' study identified the same problem. They wrote: "Not only are the engines indexing a biased sample of the Web, but new search techniques are further biasing the accessibility of information on the Web."¹⁸ They are referring specifically to search engines' measures of "popularity" to select the hits. The authors argue that:

For ranking based on popularity, we can see a trend where popular pages become more popular, while new, unlinked pages have an increasingly difficult time becoming visible in search-engine listings. This may delay or even prevent the widespread visibility of new high-quality information.¹⁹

There is another operational problem. Many databases that provide full-text articles also allow you to print them from your home or office printer. Newspaper web sites

offer archived articles for sale. When you consider that some newspapers charge anywhere from \$3 to \$5 for each article, you can see the ramifications upon a professor's salary. One freelance writer had to pay \$18 for one article. As he wrote: "Information is a national asset. There should be reasonable fees for online searches."²⁰

My fear is that this will have a chilling effect on researchers' decisions as to what database they use. It could even affect the research question itself. Let's see, choose a research question where I don't have to search X database. Forget searching this newspaper or that one. Instead, choose newspapers that have easy access. It's clear that the archives on newspaper Web sites, like the Lexis-Nexis Academic Universe database, are not designed for scholars but for casual seekers.

There is little doubt that these "mega-databases" are growing, as is Internet research usage. And the information we find—or don't find—affects our research. In the words of one journalist:

The web is transforming culture. It is transforming language, transforming information and we're seeing this in very dramatic and measurable ways, which some liken to the invention of movable type.²¹

But it would be a mistake to slavishly accept this marvel without critically examining its flaws and problems, especially as it relates to research.

The Web-based databases are great, if they can be trusted, but at this point I'm a trifle skeptical. I am uncertain about what information I have been allowed to access. The dependence on database research findings means we depend on the decisions of others who may not be familiar with the necessary rigors of scientific methodology.

The system puts an incredible amount of power in the hands of a few—the people who design the search engines, the companies who sell databases, and the newspapers and periodicals that decide which articles should be available online and for which audience. To paraphrase Lasswell: Who sells what to

whom for which audience?

In an article in *The Atlantic Monthly*, Peter F. Drucker wrote: "What we call the Information Revolution is really a Knowledge Revolution. . . . The key is not electronics. It is cognitive science."²²

We can discuss search engines and databases all we want, but what we are really talking about here is accessing knowledge.

The implications of my "research about how I can do research" seem to me to be far-reaching. The concerns extend beyond my specific difficulties in doing journalism research. They strike at the heart of all scholarly pursuits. In the 21st Century, the question is no longer only "How do we know what we know?" but "Who decides what we are permitted to know? ■

Endnotes:

¹Mick O'Leary, "Big Databases Pose Big Questions," *Online Magazine* (25) no. 3 (May 2001): 82.

²Several articles make this claim. For one example, see "Big Databases Pose Big Questions," where the author states that annual subscription rates for some of these databases start in the "high four figures" and go into the "five-figure range."

³Ibid.

⁴Steve Lawrence and C. Lee Giles, "Accessibility of Information on the Web," *Nature*, 8 July 1999, 107.

⁵Nora Paul, *Computer Assisted Research: A Guide to Tapping Online Information* 2d ed. The Poynter Institute for Media Studies, 1994, 5.

⁶Marc Faerber, ed. *Gale Directory of Databases*. Gale Group Inc., 2001.

⁷How our students use the Web for research is another issue that is not explored here. I chose to focus on faculty academic research.

⁸Lawrence and Giles, 108.

⁹Michael L. Kent, "Essential Tips for Searching the Web," *Public Relations Quarterly* (46), no. 1(Spring 2001): 26-30.

¹⁰"The Nation's Faculty: A Profile," *NEA Higher Education Advocate* (17), no. 2 (December 1999): 3.

¹¹Elizabeth Weise, "The Beginning (and End) of an Internet Beat," *Nieman Reports* (54), no. 4 (Winter 2000): 9-11.

¹²Carol Tenopir, Gayle S. Robinson Baker and William Chandler. "Getting What You Pay For?" *Library Journal* (126), no. 9 (15 May 2001): 45-58.

¹³When I was doing research at the Kettering Foundation, the foundation was using the 4.0 Lexis-Nexis software program. Just before I left, the foundation switched to the Web-based Lexis-Nexis Universe database. There is now an updated software version of Lexis-Nexis. Although there are some differences between the software and the Web-based versions, this article does not address the pros and cons of either.

¹⁴O'Leary, "Big Databases Pose Big Problems."

¹⁵Doug Isenberg, "The Database Debate," *Internet World* (7), no. 8 (15 April 2001): 25.

¹⁶The information companies that sell databases offer a variety of contracts, different plan choices with different pricing for clients. The many and varied options, how they work, their costs, and the databases received under these different contracts are not examined in this article.

¹⁷In this article I am referring only to newspaper searches. The purpose of this article is to present my personal experience as a way to explore the problems and pitfalls of database and web-based research. It is not my intent to criticize Lexis-Nexis or any other company. In fact, company representatives were extremely helpful to me as I was trying to understand my research access difficulty. I discuss

Lexis-Nexis only because this happened to be the database I was using. Indeed, it is my understanding, after talking with several librarians, that the Lexis-Nexis Academic Universe offers numerous advantages over earlier versions, providing access to many other records, other than newspapers, that I do not discuss in this article. I'm told that, for undergraduate research, it provides a great many resources for students.

¹⁸Lawrence and Giles, 109.

¹⁹Ibid.

²⁰John C. Dvorak, "Tracking Down Information Online Shouldn't Cost an Arm and a Leg," *PC/Computing* (5), no. 9 (September 1992): 44.

²¹Jon Katz, "The Future is the Net: News Online is Here to Stay," *Media Studies Journal* (Spring/Summer 1999): 14.

²²Peter F. Drucker, "Beyond the Information Revolution," *The Atlantic Monthly* (October 1999): 57.

Works Cited:

Drucker, Peter F. "Beyond the Information Revolution." *The Atlantic Monthly*, October 1999.

Dvorak, John C. "Tracking Down Information Online Shouldn't Cost an Arm and a Leg." *PC/Computing*, September 1992.

Faerber, Marc. Ed. *Gale Directory of Databases*. Gale Group Inc., 2001.

Isenberg, Doug. "The Database Debate." *Internet World*, 15 April 2001.

Katz, Jon. "The Future is the Net: News Online is Here to Stay." *Media Studies Journal*, (Spring/Summer 1999).

Kent, Michael L. "Essential Tips for Searching the Web." *Public Relations Quarterly*, (46) no. 1 (Spring 2001).

Lawrence, Steve and Giles, C. Lee. "Accessibility of Information on the Web." *Nature* (8 July 1999).

Paul, Nora. *Computer Assisted Research: A Guide to Tapping Online Information*. 2d ed. The Poynter Institute for Media Studies, 1994.

"The Nation's Faculty: A Profile." *NEA Higher Education Advocate*, 17, no. 2. (December 1999).

Tenopir, Carol, Gayle S. Robinson Baker, and William Chandler. "Getting What You Pay For?" *Library Journal* (125) no. 9 (15 May 2001).

Weise, Elizabeth. "The Beginning (and End) of an Internet Beat." *Nieman Reports* (54) no. 4 (Winter 2000).