

ONLINE®

Exploring Technology & Resources for Information Professionals



JAN | FEB 2010

VOL. 34 | NO 1

Paving Paradise

DATABASE CONTENT REMOVAL
AND INFORMATION PROFESSIONALS

Text

by AMY AFFELT

This article originally appeared in
**ONLINE, Exploring Technology & Resources for
Information Professionals**

Volume 34, NO 1, Jan/Feb 2010, Pages 14–17

 Information Today, Inc.

Copyright © 2010
by Information Today, Inc.

This article is reprinted in its entirety from the January/February 2010 issue of ONLINE magazine, published by Information Today, Inc. Used with permission. All rights reserved. Individuals may download, store, and print a single copy. All commercial uses, including making printed copies for distribution in bulk at trade shows or in marketing campaigns and all commercial reprints require additional permission from the publisher. www.infotoday.com

Contents

features

14 Paving Paradise: Database Content Removal and Information Professionals

By Amy Affelt

18 Creative Commons: Copyright Tools for the 21st Century

By Laura Gordon-Murnane

22 Resources to Encourage Entrepreneurial Creativity and Innovation

By Stephen Fadel



18



32

32 Comparing Search Engines for Quick and Dirty Answers

By Cybèle Elaine Werts

36 What Customers Want From Kindle Books

By Nancy A. Allmang and Stacy M. Bruss

40 Get a Life: Comparing Online Biography Resources

By Joann M. Wleklinski

45 What Are Libraries Doing on Twitter?

By David Stuart

columns

48 InfoLit Land

Foundations of Information Literacy: Learning From Paul Zurkowski
William Badke

51 The Dollar Sign

Finance, Personally Speaking
Marydee Ojala

54 Control-Shift

Mobile Websites With Minimum Effort
Jeff Wisniewski

58 Crawford at Large

One, Two, Some, Many: Search Results and Meaning
Walt Crawford

61 Hardcopy

You Don't Look Like a Librarian: Shattering Stereotypes and Creating Positive New Images in the Internet Age

Library Mashups: Exploring New Ways to Deliver Library Data

Complete Web Monitoring

The Kovacs Guide to Electronic Library Collection Development: Essential Core Subject Collections, Selection Criteria, and Guidelines, second edition

Deborah Lynne Wiley

64 Online Spotlight

The Virtues of Finding *and* Forgetting
Mary Ellen Bates

departments

5 HomePage

Ghost Authors, Teachable Moments
Marydee Ojala

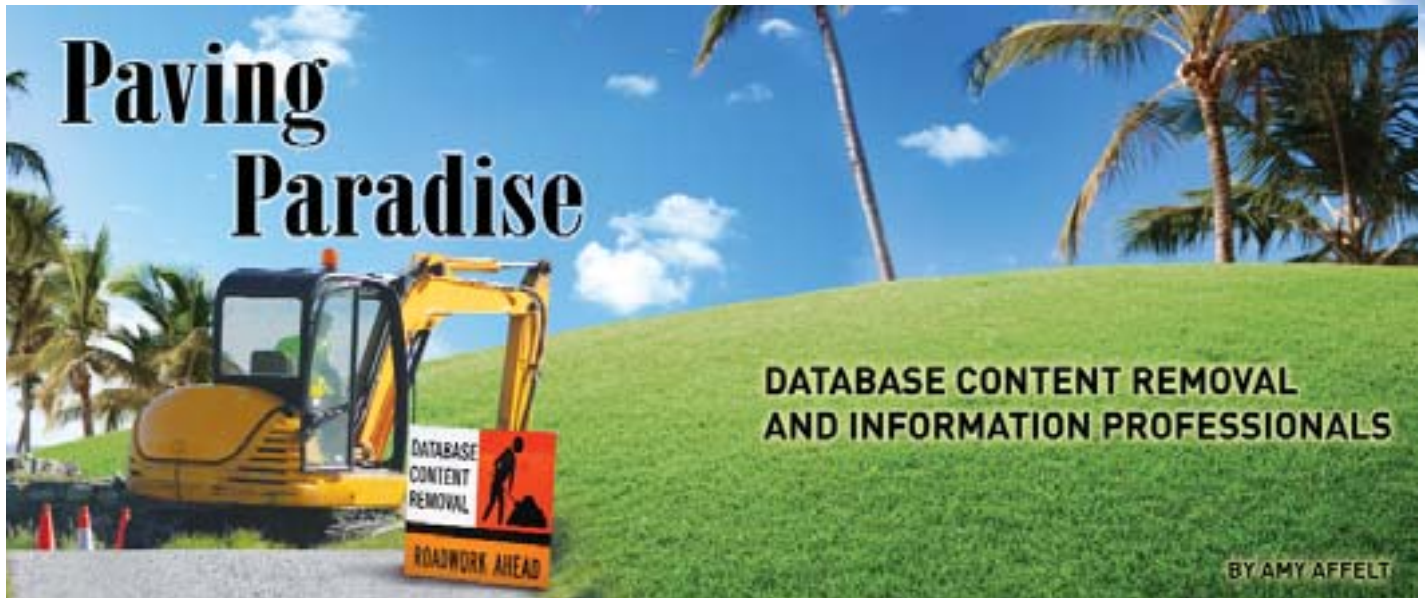
8 Industry News

Compiled by Suzanne Sabroski

13 Search Engine Update

New Search Features, Developments, and Content
Greg R. Notess

63 Index to Advertisers



Text by AMY AFFELT

I didn't expect to be channeling Joni Mitchell's "Big Yellow Taxi" as I read *The New York Times* one morning in early September. But an article mentioning a court order for Westlaw and LexisNexis to remove judicial decisions had me humming, "You don't know what you've got till it's gone."

The article discussed a lawsuit that began in 2004 (*Klein v. National Railroad Passenger Corp.*, 2:04-CV-00955, U.S. District Court, Eastern District of Pennsylvania). It involved two teenage plaintiffs who trespassed onto a parked Amtrak rail car and were severely burned when they got too close to a 12,000-volt catenary wire. The plaintiffs were awarded a jury verdict of \$24 million. While Amtrak's appeal was pending, it agreed to pay an undisclosed sum to the plaintiffs. As part of that settlement, the parties asked Judge Lawrence F. Stengel to "direct LexisNexis and Westlaw to remove the decisions and orders listed from their respective legal research services/databases." Judge Stengel agreed to do so; LexisNexis and Westlaw complied.

The legal arguments used by the attorneys to accomplish this feat remain a mystery because all of the court papers are under seal. What is known, however, is that in the 5 years of litigation, several significant legal opinions, including one discussing the burdens of proof required to demonstrate awareness of potential accidents based on previous incidents,

were handed down. Those opinions were all withdrawn. In a legal system based on precedent, the ramifications of this action are very troubling.

TAKING THE TREES

Citizens might be upset about this development not only because of an arguable First Amendment breach but also because Amtrak enjoys federal funding. Information professionals (IPs) have additional causes for concern. Previously, as an IP, the only major missing content I was aware of were articles removed as a result of the *Tasini* decision, a 2001 Supreme Court case involving freelance writers who felt unfairly compensated by publishers who reprinted their articles in electronic databases without their permission. These articles seemed to be few in number and did not seem to contain information that could make or break a case.

This may change. On Oct. 7, 2009, the U.S. Supreme Court heard oral arguments in an appeal from the Second Circuit regarding the class action settlement of publishers' payments to thousands of freelancers involving both copyrighted and unregistered works. The formal question in *Reed Elsevier, Inc., et al. v. Irvin Muchnick, et al.* is, "Does 17 U.S.C. §411(a) of the Copyright Act restrict the jurisdiction of the federal courts over copyright infringement actions?" (Updates to Literary Works in Electronic Databases Copyright Litigation 509 F.3d 136 [2d Cir. 2007] are available at www.copyrightclassaction.com.)

PUTTING THEM IN A TREE MUSEUM

In *Klein v. National Railroad*, however, the first obvious problem lies in the fact that the decisions—and, thus, the legal precedent and valuable documented research that could be used by attorneys involved in subsequent lawsuits of a similar nature—are lost. If the playing field was level and no one ever had access to these decisions, the situation would be less problematic. However, these decisions were available on both the LexisNexis and Westlaw databases at one time. Thus, many people possess PDFs of these opinions. But those conducting research after Aug. 6, 2009, are unable to obtain the PDFs through the legal online research services. They cannot fairly compete with other law or consulting firms who researched the issue earlier.

When questioned about its decision to comply with the court order, LexisNexis referred me to an outside public relations firm. I spoke with three LexisNexis employees, who offered this statement: "All LexisNexis can say at this time about the issue is that we received a letter from the U.S. District Court for the Eastern District of Pennsylvania containing an Order vacating certain Decisions and Orders issued by that Court. The letter requested and directed that those documents be removed from the lexis.com service. We have complied with the Court's request and direction. Because we cannot say more than that right now, there is unfortunately no other contact to send you to who could provide more information."



A search in genfed;mega on LexisNexis resulted in the judge's order to remove, along with a listing of the removed documents. However, I was unable to get a definitive answer from LexisNexis regarding whether or not it is the company's standard practice to make a notation for researchers to indicate when documents are removed. When doing future legal opinion searches, I am not certain whether there would be a way to tell if a search was missing a document that had been removed.

Westlaw's policy is more reassuring. Westlaw stated that "in the infrequent event that we are ordered by the court to [remove] a decision [from] Westlaw, we will comply with the order, deleting the text of the decision but keeping the title of the case and its docket number. We also publish the court's order to remove so there's a clear record of the action." When asked if content would ever be removed due to a situation other than a court order, John Shaughnessy, senior director of communications for Thomson Reuters, stated that court orders sealing a decision or draft decisions that were never formally signed or filed might also be removed. Shaughnessy stated that in all opinion removal situations, Westlaw's preferred practice is to remove only the text, leaving the header information for future reference.

SCREEN DOOR SLAMMING

Because the actions in *Klein v. National Railroad* received so much press, many bloggers, law librarians, attorneys, First Amendment activists, and others immediately uploaded PDFs of the decisions to their websites. The first two results of a Google search I conducted on Sept. 2, 2009, for "Klein v Amtrak" yielded the following sites that had links to the documents: www.volokh.com and <http://lawprofessors.typepad.com>. Furthermore, apparently the attorneys did not consider the fact that the decisions may have been printed in the hard-copy federal supplement and would be available at any law library.

IPs know, however, that most research projects we work on do not involve issues that are so widely reported. Any missing content, not just legal opinions, could cause our research abilities and credibility to be questioned. How will we know if

the vital piece of information that we need has been removed from a database? What prompts databases to remove content, and how is that removed content documented? For IPs who work in litigation, opposing counsel with documents that we thought did not exist, or were never available, is a disastrous situation that holds us directly accountable.

ANALYST REPORTS

The first pieces of content that I investigated were analyst reports from the investment banks and market research reports from firms such as International Data Corp. (IDC) and Datamonitor. Thomson Reuters is the gold standard for these reports, and John Webber, director of research, stated that "very rarely" are reports removed from Thomson Research. Thomson Reuters views itself as the publisher of these reports; the actual content of the reports is considered the intellectual property of the firms that write them. Each contributing firm has an individual contract with Thomson Reuters, and Webber emphasized that Thomson Reuters views its obligations to its contributors and to its clients with equal weight.

If an IP searched for a report that had been removed, the report would not be listed in the results set. However, Webber stated that if the IP called Thomson Reuters customer support, the internal system would be able to find out if something had been removed. This is good news for the IP. Although it involves a little tenacity in actually calling about something that one believes should be there but is not, there is a way to find out if something had been on Thomson Reuters at one time but was removed.

YOU DON'T KNOW WHAT YOU'VE GOT


Dow Jones Factiva is in a unique situation—it not only aggregates third-party data but is also part of a company that is a creator of content (*The Wall Street Journal*, Dow Jones Newswire, *Barron's*). The editorial practices and deletions policies of Dow Jones original content are completely separate from the Factiva database product policies. The latter has a formal legal deletions policy that is contained in its quality charter.

According to Ines Magarinos, manager of content quality for Dow Jones Factiva, third-party content is only removed "in

cases where it would be either a breach of contract or unlawful for it to remain." For content that is predominantly licensed, each license agreement has a clause regarding removal. Factiva is obligated by these agreements to remove content at the end of the license term or if the content provider notifies Factiva that it has a legitimate concern regarding the accuracy or legality of the content. Additionally, circumstances causing Factiva to be in violation of the law by retaining content—in violation of a court order to remove—would prompt Factiva to remove that content. Searchers would not be able to see if a document was removed, but Magarinos stated that Factiva customer service would be able to find that information and relay it to the user.

Dialog and EBSCO have very similar content availability policies, and both describe themselves as aggregators of licensed content. As intermediaries between publishers and libraries and researchers, they take their cues from the publishers in determining treatment of questionable content. Both Scott Bernier, senior director of marketing at EBSCO Publishing, and Libby Trudell, VP, market development at Dialog, LLC, emphasized the importance of upholding the licensing agreements with their providers.

When asked for specific anecdotes, both Bernier and Trudell mentioned *Tasini* as grounds for article removal. Both executives also mentioned particular circumstances that may be unique to their individual services. Bernier stated that at EBSCO, the most common issue affecting content availability is when a publisher decides to terminate its relationship with an aggregator (this issue is very familiar to those of us still longing to access Bloomberg and the *Financial Times* databases via LexisNexis). In the "extremely rare" instances in which an article is removed, it is impossible to determine on EBSCO's database whether that article was removed. However, Bernier stated that if the circumstance is that EBSCO does not have the right to include an article, it adds the following statement to the article's citation page: "This database normally includes full text of articles available from this publication. However, this particular article is not included at the request of the rights holder."



For IPs who work in litigation, opposing counsel with documents that we thought did not exist, or were never available, is a disastrous situation that holds us directly accountable.

Trudell mentioned the possibility of an article's subject complaining that it is defamatory. In that case, Dialog refers the matter to the publisher, requesting instruction regarding retention or removal. Trudell also discussed the economic impact of rarely used and outdated files. Low use and antiquated content could prompt Dialog to take a database offline.

UNTIL IT'S GONE

If the online information industry were looking for a company with best practices, ScienceDirect from Elsevier seems to fit the bill. Lindi Belfield, senior product manager for ScienceDirect, discussed a specific case that was the impetus that created Elsevier's vigorous review process for articles under consideration for withdrawal, retraction, or removal. Belfield explained that, in 2002, a scientific society demanded removal of an article from its *Human Immunology* journal, and Elsevier complied. Even at this initial removal stage, Elsevier posted a page in ScienceDirect stating, "This article has been removed by the Publisher."

However, members of the scientific community asked Elsevier to reconsider. They felt that, as professional colleagues, they should bear the burden of deciding whether content should be removed from peer-reviewed journals. As a result, Elsevier reviewed its removal process and decided in favor of the scientists. It put a process in place (http://libraryconnect.elsevier.com/lcn/0102/lc_010207.html) in which publishers fill out a form with a series of questions, ensuring that Elsevier's policy regarding retraction is followed. The form requires that the publishers clearly explain why the retraction is necessary. ScienceDirect then processes the form so that it knows the specific actions taken

regarding the retracted article, where it is stored, and the reasons that prompted the retraction. Belfield, along with the senior vice president general counsel and senior vice president of academic relations, checks the forms and the removal verbiage provided by the publisher. The original HTML of the article is then replaced with the text of the form on ScienceDirect. The PDF remains in ScienceDirect but with a red watermark on each page that says "RETRACTED." Elsevier also holds the publisher responsible for adding a retraction notice in a subsequent issue of the journal, using the same text as the retraction form and with a link to the retracted article.

Regarding complete physical removal, Belfield stated that it is very rare and would involve extreme situations such as an article that gives life-threatening information, such as an incorrect drug dosage, or an article involved in a serious legal matter. IPs can take comfort in the fact that, in these cases, even though the HTML and PDF are removed, they are replaced by HTML pages explaining the reasons for removal. Later, a notice is published and reciprocally linked.

LEAVE ME THE BIRDS AND THE BEES

Initially, I approached the research required for this article with a forbidding sense of gloom and doom. I was afraid that content was being removed from databases at will, with no consideration of the responsibility that IPs feel for the searches and results that they conduct and provide. After completing the interviews and investigations, however, I am optimistic and have a higher comfort level.

Almost none of the database companies discussed in this article remove content cavalierly. They have policies in place

to ensure that researchers are able to find out about content that has been removed. They hold their providers accountable through rigorous processes that force those providers to make strong cases for removal of content. Almost all of the vendors discussed in this article are committed to upholding these standards and keeping the removal request process quite rigorous. Similarly, we as IPs have a responsibility to pursue situations in which we believe content has been removed. It is critical that we are tenacious in contacting vendors' customer service units to obtain answers to content removal questions.

In the information industry's current climate, content is king. The ability to find quality content and add value to that content is what distinguishes the information professional from the Googler. It is refreshing to discover that in the area of content integrity, providing IPs with access to high-quality information and data remains a top priority of the database vendors. In an environment in which both IPs and vendors sometimes find themselves struggling to survive, it is my hope that we can use this shared dedication to superior quality content in order to unite and work together in ways that are mutually beneficial.

Amy Affelt (aaffelt@compasslexecon.com) is director of database research at Compass Lexecon. Comments? Email letters to the editor to marydee@xmission.com.

50% OFF



Personal subscription offer
for *ONLINE* magazine

50% savings
from current rates

U.S.: ~~\$129.95~~ **\$65**

Canada/Mexico: ~~\$115~~ **\$72.50**

Outside North America: ~~\$172~~ **\$86**

Subscribe Now!

Call 800-300-9868
to order.

ONLINE

 Information Today, Inc. 143 Old Marlton Pike • Medford, NJ 08055

www.infotoday.com/ONLINE