



University of Tennessee, Knoxville
**TRACE: Tennessee Research and Creative
Exchange**

School of Information Sciences -- Faculty
Publications and Other Works

School of Information Sciences

4-1-2005

Duplication is Ubiquitous

Carol Tenopir
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tenopir, Carol, "Duplication is Ubiquitous" (2005). *School of Information Sciences -- Faculty Publications and Other Works*.

https://trace.tennessee.edu/utk_infosciepubs/173

This Article is brought to you for free and open access by the School of Information Sciences at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.



ONLINE DATABASES

Duplication Is Ubiquitous

By Carol Tenopir

PHIL DAVIS, LIFE SCIENCES BIBLIographer at Cornell University and a 2004 *LJ* Mover & Shaker, is shaking up the scholarly community with his discovery of duplicate articles in Emerald/MCB University Press journals. According to Davis, he has found hundreds of examples of the same article published in more than one journal in at least 73 Emerald/MCB journals over 30 years. First announced at the Charleston Serials Conference in November 2004, his findings have been reported in *LJ* and its *Academic NewsWire* and in the *Chronicle of Higher Education*, with more to come (<http://people.cornell.edu/pages/pmd8>). Davis tells me, "whatever the number, no amount of premeditated covert article duplication is acceptable" and calls this "the very worst of academic publishing."

In response to Davis's allegations, Emerald undertook its own study and identified 560 republished papers from 1989 to 2004, about 1.1 percent of its total database content. Nearly all (87 percent) of the republication took place prior to 1999 and two-thirds prior to 1997. Emerald promises it is "updating the database to ensure all attributions are fully visible [including] notification of subsequent publication as well as first publication" and has addressed in-depth Davis's findings and allegations (www.emeraldinsight.com/rpsv/news/press/dual2005.htm).

Online duplication

Libraries purchase packages from multiple aggregators and publishers, which include articles from many duplicate journal titles. When *Fulltext Sources Online* (biannual, Information Today) first appeared, it averaged about two sources for each periodical title listed. Today that number is over seven. They don't all have the same date coverage and not all seven cover every journal cover-to-cover, but many of the same articles are available on multiple systems.

Most aggregators publish their journal coverage list, so librarians know they are buying overlapping titles. The special features and unique titles available in each collection drive purchase decisions. Metasearch engines and complex interlinked library systems highlight duplication and lead to the need for link resolvers that can identify duplicate sources.

Link resolvers to the rescue

Any librarian who works with link resolvers knows how many duplicate articles there are in library electronic collections. Link resolvers can either present all identifiable duplicate sources for an arti-

Duplication is a fact of library life. We need to make sure we don't pay multiple times for the same article

cle, letting users pick the one they want, or create a hierarchy of sources behind the scenes, presenting users with a single choice. The hierarchy may weigh relative costs of subscriptions or desired formats (PDF over HTML, for example).

Usage reports from link resolvers show which sources and journal titles are selected most, or least, often. In the month of January 2005, for example, no University of Tennessee (UT) user clicked through or downloaded an article from the journal *American Speech*, even though we have it in four sources with varying coverage dates. With over a hundred different online sources for journal articles, it is not unusual for a research library like UT's to have the same journal title in three or four places, all accessible through complex interlinked search systems. Thanks to the OpenURL standard, our link resolver identifies all of the sources in the total collection (both subscriptions and free sources) and lets the user choose.

Different versions, same article

An imperfect aspect of duplication is different versions of the same article. Most publishers allow authors to put

preprint versions of their articles online. Authors may put a version on their own web sites, send the article to a journal for peer review, or deposit it in a preprint server such as arXiv.org. Once the article is published, a final text version may be posted in an institutional repository with the official published version in PDF (and perhaps also HTML) on the publisher's e-journals system.

Of course, the final article may also be part of full-text third-party systems, such as ProQuest or Factiva, in plain text, PDF, HTML, or multiple formats. Even these final, edited versions may differ, since some omit copyrighted pictures,

complicated graphs, or other features. All of the various versions may end up being used and cited, especially if preedited versions are freely available while the final version is behind a paywall.

Google Scholar illuminates

Google Scholar is good at uncovering multiple versions of the same (or nearly the same) scholarly article or just multiple places for the same version. For example, for one astronomy article on "brown dwarfs," Google Scholar offered me the choice of the final published version in *Astrophysical Journal* on the publisher's password-protected e-journal system, a free PDF e-print from arXiv.org, and a PDF on an author's web site.

Duplication is a fact of library life. Metasearch engines, link resolvers, and Google Scholar are revealing multiple sources and sometimes multiple versions. Technology helps us uncover multiple versions. The next step is to make sure we don't pay multiple times for the same article.

Carol Tenopir (ctenopir@utk.edu) is Professor at the School of Information Sciences, University of Tennessee, Knoxville