Disclaimer: This is a machine generated PDF of selected content from our databases. This functionality is provided solely for your convenience and is in no way intended to replace original scanned PDF. Neither Cengage Learning nor its licensors make any representations or warranties with respect to the machine generated PDF. The PDF is automatically generated "AS IS" and "AS AVAILABLE" and are not retained in our systems. CENGAGE LEARNING AND ITS LICENSORS SPECIFICALLY DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING WITHOUT LIMITATION, ANY WARRANTIES FOR AVAILABILITY, ACCURACY, TIMELINESS, COMPLETENESS, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Your use of the machine generated PDF is subject to all use restrictions contained in The Cengage Learning Subscription and License Agreement and/or the Gale Academic OneFile Terms and Conditions and by using the machine generated PDF functionality you agree to forgo any and all claims against Cengage Learning or its licensors for your use of the machine generated PDF functionality and any output derived therefrom.

Recapturing the Past Online

Author: CAROL TENOPIR
Date: Feb. 1, 1999
From: Library Journal(Vol. 124, Issue 2)
Publisher: Library Journals, LLC
Document Type: Article
Length: 1,902 words

Abstract:
Libraries across the US are seeing an increased amount of automated information coming through to their schools. Some of these systems carry historical information which dates back only to the early seventies and maybe a few services offer articles written in the sixties. A move to update these files to include more valuable information to students as well as teachers are being devised by database producers.

Full Text:
ONLINE DATABASES are great for current information. They are usually updated more frequently than print or CD-ROM products and provide easy access to the literature of a decade or so. But as bibliographic and full-text online resources replace print versions, some librarians and teachers have noticed that student papers and even research articles often don't cite sources written before electronic coverage. Most online files begin only with articles from the 1970s or 1980s, although some (e.g., ERIC and Medline) go back to the mid-1960s. Only a very few (such as Dissertation Abstracts and Georef) provide access to journal or research information from the previous century. (Georef, from the American Geological Institute, for example, covers the geosciences literature of North America from 1785 to the present and geology of the rest of the world from 1933 to the present.)

The practice of relying solely on recent journal articles available via online sources raises more questions in certain academic disciplines. Clearly, humanities scholars can't ignore older materials, while business researchers may deem useless anything but the latest information. Most fields face less clear-cut choices. The importance of research papers or scientific discoveries from 50 or even 100 years ago may not be recognized until a specific research project "rediscovers" a finding.

Most database producers have emphasized expanding the number of titles available for current coverage or providing more timely updates. Retrospective conversion of older literature was seen as too costly and of limited appeal. Recently, however, a few database providers have begun to see the wisdom in going backwards in time. In 1999, they will launch some major initiatives to recapture the past.

Digging in the UMI vault

The most ambitious full-text backfile project is UMI's Digital Vault Initiative (DVI). Announced at the 1998 American Library Association annual convention, the DVI ultimately will provide access to 5.5 billion page images of older materials--the "world's largest digital collection of printed works."

To create the Digital Vault, UMI will scan its extensive microform collection, which includes books, newspapers, journals, magazines, and other materials. The microform collection was created over the past 60 years by microphotographing print materials, so the digital vault project represents the third generation (or third format) for these items.

The microform versions used to create the DVI are housed in actual vaults at UMI's corporate headquarters in Ann Arbor, MI. Each of three vaults is temperature- and humidity-controlled; air-filtering systems keep contaminants away from the film. Access is strictly regulated, and employees wear lint-free gloves so oils from their skin won't contaminate the film surface.

First, the oldest titles

Scanning began last May, and, even though staffers work around-the-clock on five digital scanners, it will take several years to complete. Appropriately, the first collection scanned and available will be UMI's first microfilm project. UMI's collection of early English literature was microfilmed beginning in 1938 from originals in the British Museum and includes, according to UMI, "nearly every English-language book published from the invention of printing in 1475 to 1700." Notable titles in this collection include Chaucer's The Canterbury Tales, the King James version of the Bible, and Shakespeare's First Folio edition of 1623. The 9600 individual titles will

include approximately 22 million separate page images.

The full English-literature digital collection will be available by this summer as a web-based database. The works themselves are page images only and are not full-text searchable, but MARC bibliographic records will serve as finding aids for the image files. UMI aims to provide access to the digital collection through ProQuest Direct, its online service. Libraries that already subscribe to the microform collections will have access to the digital version at a reduced rate.

Next to be scanned will be UMI's most popular periodical and newspaper titles. The full run of Time magazine from 1923 to the present will be digitized and should be available this year on ProQuest Direct; important newspapers will follow. Like most of the UMI microfilm products, the newspaper films include the full run of a title. UMI's microfilm vaults include, for example, the New York Times from 1851 to the present, Wall Street Journal from 1889 to the present, and Chicago Tribune from 1849. Since 1997, UMI has been digitizing dissertations. Older dissertations are on microfilm and may be another possible product for digitization. UMI can use subscription information to determine what titles to digitize next and survey customer response to the first digital products to determine future directions.

UMI will not abandon the present while reaching back into the past. ProQuest Direct will continue to add approximately 37 million pages of images per year of current information.

Chadwyck-Healey goes back to 1770

Chadwyck-Healey has long been known for retrospective information. Periodicals Content Index (PCI) indexes the articles in thousands of periodicals in the humanities and social sciences published since 1770. Indexing begins with the first issue of a journal and continues until 1990/91.

The complete web version of PCI includes nearly nine million articles and can be accessed by users in libraries with a subscription to PCI Web Complete. (Users can select a less costly version that indexes only from 1960/61 to 1990/91.) Besides searching for individual articles, users can request a table of contents for any journal issue. The journals in PCI are selected for their research value in consultation with librarians and from major serials bibliographies and union lists. Journals in English, German, Italian, French, Spanish, and other Western languages are included; Chadwyck-Healey claims "every effort is made to provide a complete record of publication for each journal." Eventually PCI will include 15 million records from 3500 journals. (A complete journal title list is available at <www.chadwyck.com>).

Chadwyck-Healey also publishes retrospective newspaper indexes, including The Official Index to the Times (London), 1906-1980, and Palmer's Index to the Times, 1790-1905. ArchivesUSA indexes nearly 100,000 manuscript and other special collections in U.S. libraries and manuscript repositories.

Like UMI, Chadwyck-Healey is adding archival full texts. The Digital National Security Archive provides access to U.S. government declassified documents collected from Freedom of Information Act inquiries by the not-for-profit National Security Archive. For an annual subscription fee, Chadwyck-Healey provides web access to collections by topic. The 12 topic collections made available in the fall of 1998 include documents about the Berlin Crisis, the Cuban Missile Crisis, the Iran-Contra Affair, and other 20th-century American historical events.

H.W. Wilson converts art

The H.W. Wilson Company has also begun to convert its rich resource of bibliographic information. Wilson's Art Index Retrospective project will convert all the index records covered in the print version of Art Index. By fall, Wilson will finish adding indexes from 1929 to 1984 to the currently available data online and on CD-ROM.

The first retrospective installment is already out and covers Art Index volumes 20-32 (1972-84). The second installment is due in October 1999 and will include volumes 1-19 (1929-71). When complete, Art Index Retrospective File will include indexing records for some 420 periodicals with approximately one million records.

The WilsonDoc document ordering option is built into Art Index Retrospective, so even obscure journals or older journals no longer widely available in libraries will be made available to Art Index Retrospective File subscribers. (The base price for a print copy of an article is $12.95 through WilsonDoc.)

To create such a retrospective product, Wilson did not merely scan old issues of the print Art Index. Each entry in each volume of the old indexes was first keyboarded in China. Then a separate work group--with five full-time and two part-time employees--creates a single "unit record" for each journal article from the multiple entries per article in the print version (e.g., subject heading and author). They also correct errors in entries.

As Barbara Chen, associate director of Indexing Services for Wilson, explained to me, "In the `olden days,' when the article had three subject headings, it was typed three times. Errors crept in." The current indexing staff has attempted to identify errors, including variations in article titles. In addition, periodical title abbreviations had been used in the past (and they had changed over 50 years) so full periodical titles and ISSN are added to each record.

Wilson indexes have always been known for their attention to detail and careful authority file maintenance. To create just the first installment of Art Index Retrospective, Chen said, "We had to reconcile over 99,000 personal names, 56,000 topical terms, and 4000 corporate names with subjects in our current Art subject file and our current names file." She was "amazed at how the cataloging

rules changed over 50 years." Wilson expects many more names will need to be reconciled in the second installment.

It would be impossible, even for Wilson indexers, to reconcile every name or subject that fails validation. Certain types of names can be reconciled more easily than others. For example, names of all popular artists and all names with five or more postings are reviewed. All names that vary by only one character are made consistent (for example, Pieter vs. Peter); indexers use the Soundex computer program to identify names with minor spelling variations.

The current Art Index subject headings list serves as the basis for the retrospective product's subject authority file. According to Chen, "We tried to take the older article terms and make them correspond to the terms we are presently using. It is not always possible, because we don't have the actual articles in front of us." For example, the older subject term "Hotels, Restaurants, etc." was replaced by separate terms "Hotels" and "Restaurants." Unless the article title offers a clue, indexers can't tell which of the two new terms best applies. Cross references for presently used terms are added to the retrospective product to help users move between new and older terms.

While Wilson plans to introduce only Art Index Retrospective File as a backfile product this year, it will consider other projects. "We have learned a lot in the first six months and we will learn more in the next," said Chen. Wilson thinks that retrospective products have a great deal of potential, but, Chen added, "we'll have to see how well this one sells."

Other ventures into the past

The American Psychological Association has recently also gone back in time with indexing. Historic PsycINFO is a recent addition to PsycINFO subscribers on SilverPlatter. Historic PsycINFO takes the indexing of psychology journals back to 1887-1966, when the field was first beginning, and includes over 284,000 records. If these early retrospective products are successful, expect more backfiles. Older indexing and full texts are important to scholars and to posterity, but these new products from information companies must be commercially successful to justify others.

If librarians believe backfiles are important, they must demonstrate this priority by committing resources. When budgets are tight, going back in time may be feasible only in research libraries. How much are you willing to pay to access the literature of the past?

Carol Tenopir is Professor at the School of Library and Information Science, University of Tennessee at Knoxville. Her E-mail address is tenopir@utkux.utk.edu

Copyright: COPYRIGHT 1999 Library Journals, LLC. A wholly owned subsidiary of Media Source, Inc. No redistribution permitted.
http://www.libraryjournal.com.proxy.lib.utk.edu:90/
Source Citation (MLA 8th Edition)
TENOPIR, CAROL. "Recapturing the Past Online." *Library Journal*, 1 Feb. 1999, p. 39. *Gale Academic Onefile*, https://link-gale-com.proxy.lib.utk.edu/apps/doc/A53728601/AONE?u=tel_a_utl&sid=AONE&xid=60422c6b. Accessed 20 Dec. 2019.
Gale Document Number: GALE|A53728601