

Oldies but Goodies

COMMON WISDOM says that currency is what is valued in online databases. With the exception of genealogists and historians, it is assumed that readers don't want old materials. From the publisher's side, it just isn't cost effective to convert older bibliographic or full-text articles to digital form. For those who want older material, print—at least until recently—was thought to be the medium of choice.

This common wisdom is being proved not so wise, as readers are accessing both new and older information when it is available online, and publishers are rushing to digitize long runs of magazine and journal articles and the indexing tools that point to them.

ProQuest digitizes newspapers

From its prior incarnation as University Microfilms, Inc. (UMI), ProQuest has a history of respect for old materials. It's been five years since ProQuest announced its Digital Vault Initiative, a project to convert to digital form much of its extensive historical collection of newspapers, dissertations, and other materials. That project was followed by the ProQuest Historical Newspapers project, with the ambitious goal of digitizing "every back issue of America's most respected newspapers."

This summer we began to see results when the digitization of all back files of the *Wall Street Journal* and the *New York Times* was completed. The *NYT* historical file covers from 1851 through 1999 and includes nearly three million pages with over 25 million articles. The *WSJ* extends from 1889 through 1985 and includes more than one-half-million pages and five million articles. Next up are the *Washington*

Post (from 1877 to the present) and *Christian Science Monitor* (1908–90), both of which should be completed by spring 2003. Look for complete back files of the *Chicago Tribune* and *Los Angeles Times* by the end of 2004.

Completely searchable

What is unique about ProQuest's historical runs is that they combine a page-by-page scan of the papers with an article-by-article search capability. This is the result of a complicated process involving scanning microfilm to create full-page image files. Individual articles on each full page are then zoned, i.e., individually identified, threaded (meaning that if an article begins on one page and continues to another, the images will be presented as a single image), and linked to their appropriate page. All are then run through optical character recognition software to create searchable ASCII text, the images are enhanced, and each article is tagged with XML elements.

This is done not only for substantive articles but for every individual object on a newspaper page. ProQuest's list of article identifiers shows the range of items that can be searched and viewed. These include articles, obituaries, advertisements, comics, wedding announcements, birth notices, weather reports, letters-to-the editor, editorials—virtually everything in the paper.

The product manager said that ProQuest Historical Newspapers "finds a lot of use among genealogists researching their ancestors." It could also be used by advertising agencies, students in American studies or popular culture, journalists researching background for a story, or senior citizens on a nostalgia trip. Eliminating microfilm and separate newspaper indexes will be welcomed by many.

Gale products coming soon

Gale is also busy digitizing old materials. It will digitize *The Times* (London) all the way back to 1785 and the papers of Winston Churchill. Books are next in line for Gale. The company re-

cently announced a cooperative venture with the British Library, other research libraries, and the English Short Title Catalogue Committee to create a massive historical digital project called The Eighteenth Century—Complete Digital Edition. Eventually this product will include about 150,000 English-language books published between 1701 and 1800 and total approximately 20 million pages.

From mid-2003 through 2006, titles will be made available in subject category collections. The first will be history and geography, followed by literature and language; social science; religion and philosophy; science, technology, and medicine; law; fine arts; and reference. Each will be created by scanning the microfilm product, creating searchable records, and adding MARC records. Metadata will be added, including the title and contents pages, and all illustrations will be available. This combination of images with fully searchable texts and bibliographic information sets these commercial products apart from many free web sites of historical materials.

Chemical Abstracts links to old

At this summer's 2002 meeting of the American Chemical Society (ACS), ACS and Chemical Abstracts Society announced that old Chemical Abstracts records now link to full-text articles from the American Chemical Society Journal Articles archive. In 2001, Chemical Abstracts Society abstracts were brought online all the way back to its 1907 beginning. At the same time, American Chemical Society began digitizing its journal archives back to 1879.

If a search on the back file of Chemical Abstracts turns up an interesting old article that was published in an American Chemical Society journal, users can now use the ChemPort feature to link to the full text. ChemPort works with all Chemical Abstracts Society search systems, including STN Express, STN Easy, STN on the Web, SciFinder, and SciFinder Scholar. Of course, if an article indexed in Chemical Abstracts



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

ONLINE DATABASES

was in a journal published somewhere else, users will need to track down that copy elsewhere.

JSTOR leads the way

JSTOR is one of the pioneers in digitizing complete back runs of scholarly journals. Its success demonstrates the value of older journals to many different types of users. Started as a research project sponsored by the Andrew W. Mellon Foundation at the University of Michigan, JSTOR's original impetus was to help libraries deal with the storage and preservation problems posed by paper journals. It became an independent, not-for-profit organization in 1995 and has grown steadily since.

Currently, JSTOR provides access to page image and searchable text files for over 240 journals. (This translates to over 1.5 million individual articles online.) Many are in the social sciences and humanities, but JSTOR includes a variety of topics.

Journal titles belong to collections, including Arts & Sciences, General Science, Ecology & Botany, and Business. The newest collection is Language & Literature, out this fall. The Arts & Sciences II Collection, announced in 2001 and which is being released as titles are completed, will have a full complement of more than 100 journals by the end of the year. More collections are under development, including Art History, Music, Education, and Law.

JSTOR digitizes print journals from approximately 176 publishers and has begun work on adding born-digital journals as well. New titles are added continually as time and resources permit. (The JSTOR web site lists over 100 journal titles that have agreed to participate in the near future.) To select titles, JSTOR looks for journals that have a substantial back file, are requested by researchers in that field, have a relatively high library subscription base, and are highly ranked in the ISI Journals citation reports.

Studying usage

There is no question that JSTOR journals are being used. Over 1400 institutions participate in JSTOR access, including more than 950 U.S. and close to 460 international institutions. A majority of these are colleges and universities, but many are government agencies. Some 6.2 million articles were printed from nearly 10.9 million browses of ci-

tations, titles, or tables of contents from January to early September 2002. The 2002 use will significantly exceed 2001's totals, when about 6.3 million articles were printed.

In "Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR," Kevin Guthrie, president of JSTOR, stated that the av-

Some 6.2 million
JSTOR articles were
printed from nearly
10.9 million browses
of citations, titles,
or tables of contents
from January to early
September 2002

erage age of downloaded articles varies with the subject discipline. The average age of articles requested in economics is 13 years old while mathematics is as high as 32 years old. Highly requested articles from JSTOR are not necessarily those that get the most citations, however, showing that there are many reasons why someone reads a journal article and that citations alone are not enough of an indicator of an article's significance.

Two years ago JSTOR commissioned a survey of faculty members in U.S. colleges and universities to find out how they use electronic resources and their libraries. The survey, Faculty Response and Attitudes Toward Electronic Resources, revealed that over 60 percent of faculty are comfortable using electronic resources and consider them important to their research. A majority (52 percent) of economists believe that in the future they will be able to conduct their research without setting foot in the library, but only 22 percent of humanists agree with this statement. Humanists are also much more likely to see the need for hard-copy archives in the future.

Data on reading agree

JSTOR's data that indicate the heavy usage of older articles are con-

sistent with reading data I have collected with Donald W. King over the years. We have found that scientists, social scientists, and engineers vary in their reading of older articles, but, on average, about one-third of readings are to articles more than one year old. We asked astronomers in a survey last year about the last article they had read. The date of the last article read varied from 1938 to 2002.

A large proportion of the readings were current: 64 percent were from 2001 or 2002, nearly 74 percent were from 2000 to 2002, and 93 percent were from articles published within the last ten years. Still, as our previous studies have shown, the older readings are rated by respondents to be of high value and are designated as important.

Fond memories

Often an older article is a rereading, perhaps something the researcher remembered from when it was new and returned to when the need arises. Other times users are doing research that requires a long-term perspective or are starting an interdisciplinary study that demands review of the past literature.

When older materials are digitized and linked to popular indexing and abstracting tools, we can expect them to get at least as much use in many science fields and much more use in disciplines such as humanities or mathematics. Although neither JSTOR nor I used the term "goodies" in our surveys, clearly older materials are just that—oldies but goodies for many readers.

LINK LIST

Chemical Abstracts Society

www.cas.com

The Eighteenth Century—Complete Digital Edition

www.gale.com

Faculty Response and Attitudes Toward Electronic Resources

www.jstor.org/news/2002.03/SurveyMarch2002.html

JSTOR

www.jstor.org

ProQuest Historical Newspapers

www.proquest.com

Revitalizing Older Published Literature

www.jstor.org/about/preliminarylessons.html