PERSPECTIVES ON TECHNOLOGY

HISTORIC JEWISH PERIODICALS/ NEWSPAPERS AND THE WEB

Heidi Lerner

Scholars in all areas of Jewish studies depend on historic periodicals and newspapers to conduct their research. We are all too familiar with the many difficulties in accessing and using these materials: individual titles and collections are dispersed across continents; copies, due to the high acid content of the paper on which they were printed, are fragile, hard to handle, and difficult or prohibitively expensive to maintain. While formats such as microfiche or microfilms solve the problem of deterioration, they do nothing to facilitate access, nor do they lend themselves to easy browsing or searching.
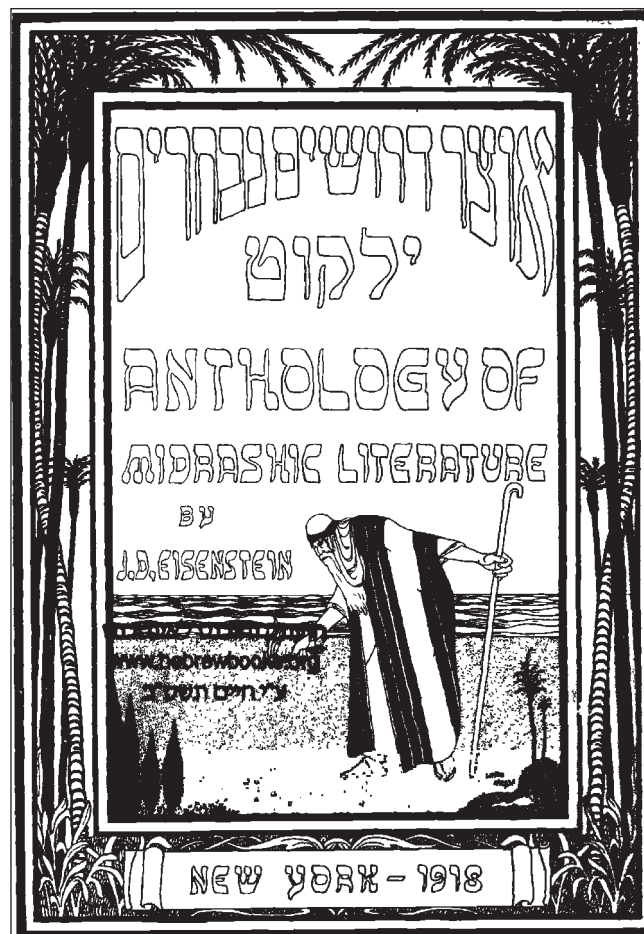
Information technology offers solutions to many of these problems. Digitization and Web access can make these texts more widely available, less subject to deterioration from handling of the physical artifacts (although there is the issue of media obsolescence to be concerned about), and more easily searched and browsed. Printed, manuscript, and pictorial information is converted into digital image files for use in computer-based applications. This is done in several ways: most commonly, a

Otzar Drushim Nivcharim by Rabbi J.D. Eisenstein. New York, NY, 1918. Courtesy of www.hebrewbooks.org.

scanner is used to create an electronic image of a document much like a photocopy machine does, except that the image is viewed on a computer screen. Sometimes a digital camera is used to create computer records of images instead of recording them on film. Ideally, the image files will provide faithful replications of the original documents. The integrity of the original document needs to be retained, and different processes and standards are developed for creating both high-resolution archival and smaller deliverable image files.

The drawback of simply using a scanner or digital camera to reproduce text is that these only produce images of the page. The text can be read, but not searched. To solve this problem, text can be rekeyed, a process that is time-consuming and costly, ideally involving not only someone to input the text, but also an editor and/or proofreader. Optical Character Recognition (OCR) software enables the computer to recognize characters as it scans the text line by line, by attempting to match character images into predefined patterns and lexical contexts, then to represent each with the appropriate character code. The advantage of using OCR software is that text can then be searched. Complex programming, however, is required to plot formal varieties of characters, as well as to integrate some sort of lexical database (such as automated spelling or grammar checkers) to increase the accuracy of simple shape recognition. In the case of Hebrew, functional OCR software would be able to recognize and distinguish the variety of printed typefaces that can simultaneously exist on a single page (such as the common juxtaposition of square and semicursive "Rashi" characters) and, when necessary, support vocalization and diacritics. Unfortunately, no commercial or public program that satisfies these criteria is available. Even more problematic are multilingual texts (lack of adequate computational linguistic support) and manuscript documents.

In recent years, an increasing number of institutions, both public and private, as well as individuals are recognizing the benefits of easier access and enhanced protection that digitization offers to their

collections. They are initiating efforts to preserve historic Jewish publications and offer them electronically via the Internet to the public.

The Compact Memory project (http://www.compactmemory.rwth-aachen.de/cm2/html/),* based at institutions in Aachen, Frankfurt, and Cologne, provides free, full text access to some of the major nineteenth- and early twentieth-century German-language, Jewish periodicals. The images include illustrations and small advertisements, as well as the text of the articles. These periodicals cover the religious, social, political, and cultural aspects of Jewish life in German-speaking Western and Central Europe. Access to these periodicals has been extremely limited due to the vast and systematic destruction during the Nazi period, which has resulted in surviving paper copy being scattered in just a few libraries.

The project, initiated in 2000, is scheduled for completion in 2006. Text searching of those titles is not yet available, but all the issues of the following periodicals have been scanned as images: *Allgemeine jüdische Wochenzeitung, Allgemeine Zeitung des Judentums, Altneuland, CV-Zeitung, Der Jude, Der Morgen, Der Orient, Die Freistatt, Die Welt, Esra, Im deutschen Reich, Jahrbuch fur jüdische Geschichte und Literatur, Jeschurun, Menorah, Mitteilungen des Gesamtarchivs der deutschen Juden, Neue jüdische Monatshefte, Ost und West, Palästina, Wissenschaftliche Zeitschrift fur jüdische Techmologie, Zeitschrift fur Demographie und Statistik,* and *Zeitschrfit für die Geschichte der Juden in Deutschland.*

The Laura Schwarz-Kipp Institute for Advanced Technology in the Humanities at Tel Aviv University, under the direction of

Dr. Ronald Z. Zweig, has brought to the Web the entire run of the *Palestine Post* (1932–1950). This amounts to over 40,000 pages of broadsheet newsprint (http://kipp.tau.ac.il/Archive/skins/Palestine/navigator.asp). The technology developed for the project enables full-text searching and provides high-resolution images.

The Society for the Preservation of Hebrew Books has brought more than 360 American Jewish journals online (www.hebrewbooks.org).

The Society for the Preservation of Hebrew Books has brought more than 360 American Jewish journals online (www.hebrewbooks.org). Among the scanned Hebrew journals available for reading and browsing, printing and downloading are: *ha-Pardes, ha-Mesiloh, ha-Tevunah, ha-Yehudi, Kol Yerushalayim, ha-Keri'ah veha-Kedushah, ha-Kokhav, Kerem asefat ḥakhamim, Bet Yitsḥak, Degel Yisra'el, Idno, Ohel Yosef, Or ha-Me'ir, Or ha-mizrah, Talpiot, ha-Mitspeh, Kol Ya'akov,* and *Kol Yerushalayim.*

Some humanities online databases are offering retrospective full-text access to journals relevant to Jewish studies. From within American Theological Library Association Serials (ATLAS), selections currently include the *Hebrew Union College Annual* (1949–1996), *Journal for the Study of the Old Testament* (1976–2001), and *Biblical Archaeologist* (1949–1997). JSTOR offers among its titles, the *Journal of Near Eastern Studies* (1942–1997), *American Journal of Semitic Languages* (1894–1941), and *Hebraica* (1884–1895). Periodical Contents Index (PCI), while still in the early stages of introducing full-text access, offers

full retrospective indexing of fifty-seven periodicals in Jewish studies, including *Jewish Quarterly Review, American Jewish History, Estudios sefardies, Historica Judaica,* and *Jewish Historical Studies,* etc. Access to these databases is by subscription only (site-license, IP-recognition basis).

Hopefully, this brief overview of digitization and the historic Jewish press will stimulate an interest in further explorations of how computers and electronic information can impact our own research and scholarly output. The editors invite you to comment, contribute, or suggest other topics for this newsletter relating information technology to our field of Jewish Studies by contacting Heidi Lerner at *lerner@stanford.edu.*

Resources:
1. *Guides to Quality in Visual Resource Imaging* (from the Research Libraries Group: http://www.rlg.org/visguides/)
2. *RLG DigiNews* (From the Research Libraries Group: http://www.rlg.org/preserv/diginews)
3. Digital Library Federation homepage (http://www.diglib.org/publications.htm)
4. Zweig, Ronald W. "Lessons from the Palestine Post Project," *Literary and Linguistic Computing* 13: 2 (1998): 89–95.
* Links provided were valid as of August 2003. Due to the volatility of the Web, they may no longer work.

Heidi Lerner is the Hebraica/Judaica Cataloger at Stanford University