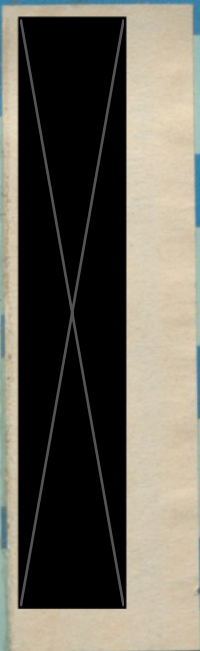
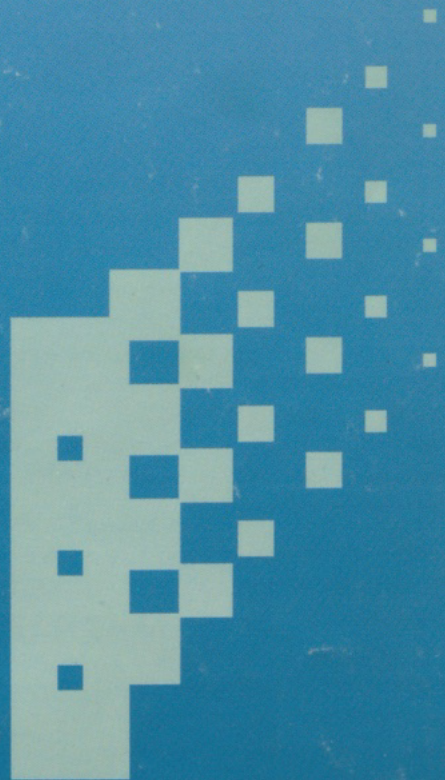
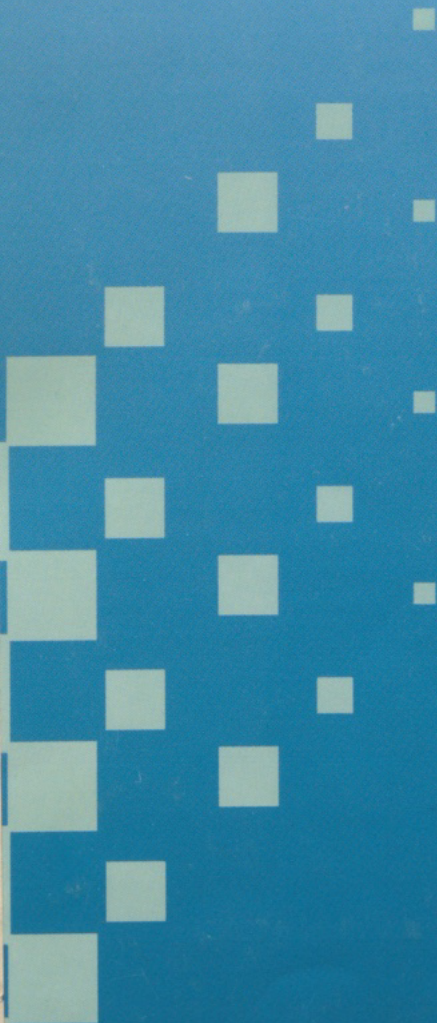
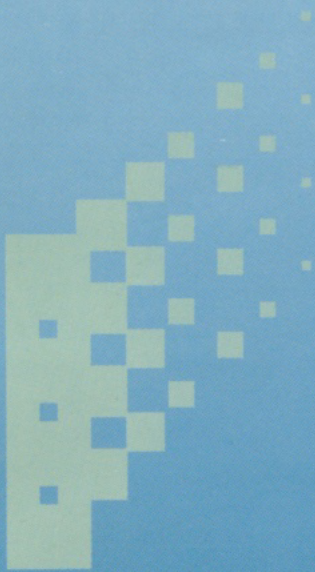


# Information Technology and Libraries

Volume 13    Number 1    ISSN 0730-9295

**March 1994**



# The 95% Solution

You get one chance with authority control, so it's important to get it done right. LTI guarantees that its affordable, machine-only authority control will link 95% or more of your library's controlled headings to an LC or LTI authority record. No exceptions! No excuses!

When manual review is requested, only professional librarians are used as editors and link rates approach 100%.

LTI maintains the complete LC MARC authority files (updated weekly), supplemented with over 410,000 LTI authority records and 350,000 proprietary "cross links."

Contact LTI for more information on authority record link results.

*"Authority Control for the 21st Century"*



**LIBRARY TECHNOLOGIES, INC.**

1142E Bradfield Road Abington, PA 19001  
(215) 576-6983 Fax: (215) 576-0137  
(800) 795-9504

# WHY WE'RE GIVING YOU A BETTER WAY TO SEARCH DATABASES.

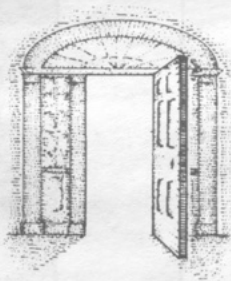
**W**e started in 1985, database searchers committed to better search software design. We became the premier vendor of Medline, then expanded our catalog to other databases. Last year we won Information World Review's PRODUCT OF THE YEAR for faster, easier search

software. But a better way means meeting the evolving needs – individual and campus wide – of today's library users.

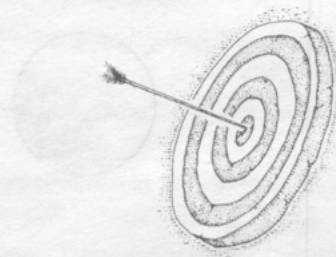
**A**nnouncing OVID: a database interface also flexible it molds itself to your search environment.



*With OVID you're free to move from one operating system to another without retraining. OVID's Common User Interface assures identical functionality in DOS®, Windows™ and UNIX®.*



*A haven for beginners searching without supervision, OVID's Easy Mode has on screen prompts. The more experienced can pull-down menus showing an array of search options. Experts will feel at home using online syntax.*



*Search with natural language if you like. OVID mapping cuts through the mystery of controlled vocabularies, homing in on precisely matching subject headings.*



*There's an unprecedented array of search tools – indexes, thesauri, limits and fields – many never before available in an interface. They're all standard OVID features.*



*HELP for every search function is context-sensitive and on screen, never more than a keystroke or mouse click away.*

**OVID. A better way to search ERIC, Current Contents®, PsycINFO®, Medline, Readers' Guide Abstracts, EMBASE and more.**

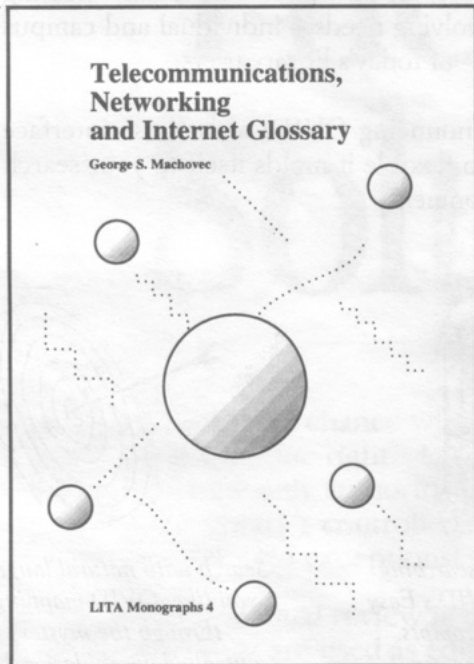
**CD PLUS Technologies**

New York 800-950-2035/212-563-3006

London 44-(0)81-748-3777

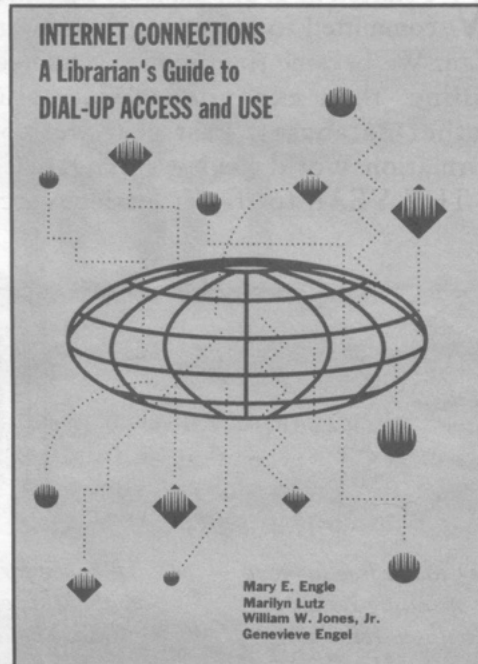
Amsterdam 020-672-0242

# KEEP CURRENT WITH TELECOMMUNICATIONS AND INTERNET



**Telecommunications, Networking and Internet Glossary**, by George S. Machovec. (LITA Monographs No. 4) 106p. pbk. ISBN 0-8389-7697-2 \$18.00

Developed to assist librarians, information managers, and students in remaining current with the constantly increasing amount of new terminology in the high-growth communications field, this new glossary from CARL (Colorado Alliance of Research Libraries) technical coordinator George S. Machovec provides hundreds of definitions to telecommunications, networking and internet terms that frequently appear in library automation technology. A handy reference for today's librarian.



**Internet Connections: A Librarian's Guide to Dial-Up Access and Use**, by Mary E. Engle, et al. 166p. pbk. ISBN 0-8389-7677-8 \$22.00

This guide is for librarians and library patrons who are not already members of the Internet family of users. It is a snapshot of the networking services available at the time of publication, a discussion of the concepts and terms in use, and a bibliography of the notable guides that document the search techniques, navigational tools, and information resources available. It is designed to help you get started right away.

**To order**, call 1-800-545-2433 and press 7. Call between 8:30 a.m.-4:30 p.m. CST.

**LITA**

Library and Information Technology Association  
a division of the  
American Library Association

Thomas W. Leonhardt, Editor  
University of Oklahoma, Norman  
Marjorie E. Bloss, Managing Editor  
Center for Research Libraries, Chicago, Illinois  
Susan B. Harrison, Book Review Editor  
New York Public Library, New York  
George S. Machovec, Software Review Editor  
CARL, Denver, Colorado  
Brian Sealy, Tutorials Editor  
University of Michigan, Ann Arbor

EDITORIAL BOARD: Joan M. Aliprand, Research Libraries Group, Mountain View, California; Rao Aluri, Boone, North Carolina; Walt Crawford, Research Libraries Group, Mountain View, California, ex officio; Patrick Flannery, Houston Academy of Medicine, Texas Medical Center Library, Houston; Ben-Ami Lipetz, School of Information Science and Policy, State University of New York, Albany; Marilyn Lutz, University of Maine, Orono; Dilys E. Morris, Iowa State University, Ames; Kurt R. Murphy, Arizona State University, Tempe; Randy L. Pedersen, University of North Dakota, Grand Forks; V. Louise Saylor, Eastern Washington State University, Cheney; Xiao-Yan Shen, University of the Pacific, Stockton, California.

*Information Technology and Libraries* (ISSN 0730-9295) is the official publication of the Library and Information Technology Association, a division of the American Library Association, 50 E. Huron St., Chicago, IL 60611; Executive Director: Linda J. Knutson. The journal is issued quarterly in March, June, September, and December.

*Information Technology and Libraries* publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 88 of the March 1992 issue. Manuscripts of articles, communications, and news items should be addressed to the editor: Thomas W. Leonhardt, *Information Technology and Libraries*, Bizzell Library, 401 W. Brooks, University of Oklahoma, Norman, OK 73019-0528; (405) 325-2611; Internet: leonhardt@aardvark.ucs.uoknor.edu. Copies of books for review should be addressed to: Susan B. Harrison, *ITAL Book Reviews*, The New York Public Library, 455 Fifth Ave., New York, NY 10016. Copies of software for review should be addressed to: George S. Machovec, *ITAL Software Reviews*, Colorado Association of Research Libraries, 3801 E. Florida Ave., Ste. 370, Denver, CO 80210-2542. ADVERTISING: contact Stuart Foster, Ad Sales Manager, and Dolores LaPointe, Ad Coordinator, c/o Choice, 100 Riverview Center, Middletown, CT 06457; (203) 347-6933.

As a matter of policy, *Information Technology and Libraries*, as the scholarly organ of LITA, does not review LITA publications. Notice of new publications from LITA will generally be found in the "Other Recent Receipts" column following reviews.

*Information Technology and Libraries* is a requisite of membership in the Library and Information Technology Association. Subscription price, \$22.50, is included in membership dues. Nonmembers may subscribe for \$50 per year in the U.S.; \$55 in Canada, Mexico, Spain, and other PUAS countries; \$60 in other foreign countries. Single copies, \$15. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

**Circulation and Production:** ALA Production Services (Eileen Mahoney; Dianne Rooney; Bruce Frausto, and

# Information Technology and Libraries

Volume 13, Number 1

March 1994

ISSN 0730-9295

## SPECIAL ISSUE: Electronic Texts: Where Next?

- 4 Editorial
- 6 Electronic Texts: Introduction KATHARINA KLEMPERER
- 7 Center for Electronic Texts in the Humanities MARIANNE I. GAUNT
- 15 "A Library and Apparatus of Every Kind": The Electronic Text Center at the University of Virginia DAVID M. SEAMAN
- 20 Building the Digital Library: The University of Michigan's UMLibText Project BETH FOREST WARNER and DAVID BARBER
- 27 The Rutgers Inventory of Machine-Readable Texts in the Humanities: Cataloging and Access ANNELIES HOOGCARSPEL
- 35 Poststructuralism and the ARTFL Database: Some Theoretical Considerations MARK WOLFF
- 43 Electronic Texts in the Humanities: A Selected Bibliography ANITA K. LOWRY
- 51 **SPECIAL SECTION: Future Possibilities in Information Technology and Access** MILTON WOLF, EDITOR
  - 51 Introduction: Science Fiction—A Mainstream Force MILTON WOLFF
  - 51 *Jurassic Park* and AI Jolson: Thinking about the Information Revolution CONNIE WILLIS
  - 53 The Good and the Bad: Outlines of Tomorrow DAVID BRIN
- 63 **Tutorial** BRIAN SEALY, EDITOR
  - 63 Mapping 1990 Census Data PATRICK YOTT
- 71 **Software Reviews** GEORGE MACHOVEC, EDITOR
- 77 **Book Reviews** SUSAN B. HARRISON, EDITOR
- 86 **News and Announcements**
- 95 **Letter**
- 95 **Index to Advertisers**

Cover Design by Jim Lange Interior Design by Dianne M. Rooney

Donavan Vicha), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by LITA or the ALA.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdelnyi Vypusk, and Science Abstracts Publications*. Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, New Search, and Social Sciences Citation Index*. Microfilm copies avail-

able to subscribers from University Microfilms, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. ∞

Copyright © 1994 American Library Association. All material in this journal subject to copyright by ALA may be photocopied for the noncommercial purpose of scientific or educational advancement granted by Sections 107 and 108 of the Copyright Revision Act of 1976. For other reprinting, photocopying, or translating, address requests to the ALA Office of Rights and Permissions.

## Editorial

You will have noticed by now that *ITAL* is a larger size than the last time you received a copy. Despite a name change and the addition of varied cover colors over the years, the size and format has not changed since *ITAL*'s beginning, when it was known as *JOLA, The Journal of Library Automation*.

Times change, and for some time now the scope of *ITAL* has changed. Automation and information technology are phrases that convey only a part of what LITA and *ITAL* are all about. The contents of this issue attest to that.

The two special sections in this issue emphasize the intellectual aspects of the work that many of us do or support, even if only indirectly. We are not machines, and the machines we work with are merely tools. We must never forget that or we have lost our purpose and will find ourselves on the outside looking in.

Special thanks are in line to Dianne Rooney of ALA Publishing Services, who worked hard to come up with the cover design (the cover artist is Jim Lange) and the new typeface and layout. We all too often take such things for granted, but they are important nonetheless. We hope that you find the new design attractive and that you find *ITAL*'s new face easier to read. We also hope that the contents are just as appealing.

Please let me hear from you about *ITAL*. It is your journal, and without loyal, caring readers, it doesn't really exist. You can reach me by post and also via the Internet at [leonhardt@aardvark.ucs.uoknor.edu](mailto:leonhardt@aardvark.ucs.uoknor.edu).—Thomas W. Leonhardt

# Now get unlimited, low-cost access to the PAIS Database via leased tape

**For as low as \$3,000 per year, you can have unlimited access to the PAIS International Database, the premier bibliographic index to public and social policy issues.**

***Economical, fixed-fee pricing.***

With leased magnetic tapes, there's no need to worry about usage and teleconnect fees. Your users can search PAIS as often as they like, for as long as they like, all for one annual rate.

***Flexibility and ease of use.*** PAIS is compatible with most OPACs. That means your users don't have to learn different search software, and *you* don't have to acquire new hardware.

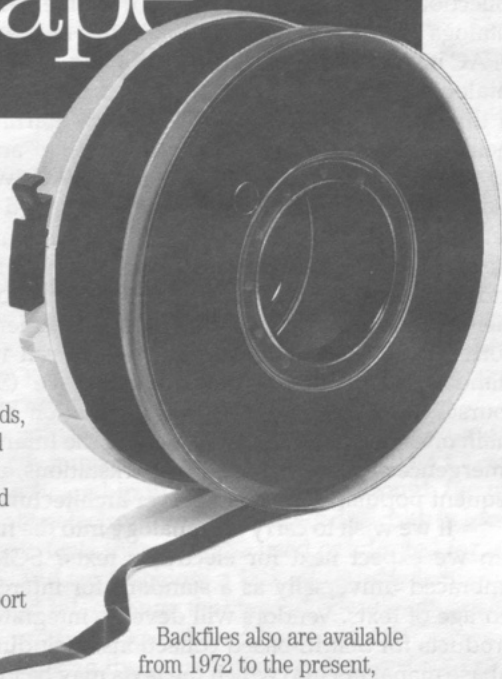
***Institution-wide availability.*** With PAIS International on your network, users in dorm rooms and offices across an entire campus or organization can access the database 24 hours a day.

***A total information resource.***

The PAIS magnetic tape includes references to over 380,000 journal articles, government documents, statistical directories, monographs, conference reports and more. The database also references literature published worldwide in English, French, German, Italian, Spanish and Portuguese. Bibliographic records, a Thesaurus File, and Periodical and Publisher Directories round out this unique resource. ISSN, ISBN, LC and Superintendent of Documents are included in the data.

***Full telephone support.*** PAIS support is never more than a fast phone call away.

An annual subscription to PAIS on magnetic tape includes 12 monthly updates and complete documentation.



Backfiles also are available from 1972 to the present, either as a whole file or in five-year units. For more information or a free sample tape, use the coupon below or call toll-free, 800-288-PAIS.

Please send me more information on PAIS Database tapes.

Please call me about receiving a sample tape and tape documentation.

Name \_\_\_\_\_

Institution \_\_\_\_\_

Address \_\_\_\_\_

City/State/ZIP \_\_\_\_\_

Phone \_\_\_\_\_ Fax \_\_\_\_\_

**PAIS<sup>®</sup>** No one  
looks at the world  
like PAIS

**Public Affairs Information Service<sup>®</sup>, Inc.**

521 West 43rd Street, New York, NY 10036-4396  
800-288-PAIS, 212-736-6629 (in NYC)  
Fax: 212-643-2848

Editing this special section on scholarly electronic texts has been an educational experience. Until now it has been difficult to find an assemblage of articles that gives an overview of the state of the art in electronic text resources, especially in a library-oriented journal. This collection should be especially useful because it provides many pointers to further sources of information, both printed and available electronically on the Internet. Thanks are due to the authors who contributed this excellent set of articles.

In many ways, the state of electronic scholarly text collections is analogous to that of online public access catalogs (OPACs) in the early 1980s. Just as the pioneering OPAC institutions were then starting to convert their card catalogs, the electronic text pioneers, from whom we hear in this special section of *ITAL*, are now starting to build their own text collections, often "converting" and marking up the texts on their own. The early OPACs were home-grown because there were no vendor-supplied access and delivery products, much less integrated systems; similarly the current text centers are using a variety of search engines and delivery mechanisms and supplying a good deal of their own access software. The OPAC movement was supported by the MARC standard; e-texts can rely on the Standard Generalized Markup Language (SGML). Of course, differences in the general situation abound, the main ones being the pervasiveness of the Internet and the emergence of powerful desktop workstations, and the consequent popularity of client-server architectures.

If we wish to carry this analogy into the future, what can we expect next for electronic texts? SGML will be embraced universally as a standard for interchange and storage of texts. Vendors will develop integrated turnkey products for SGML-based collections, including both database managers into which the texts may be loaded and a variety of user-interface clients. Perhaps these will be included in existing integrated catalog systems, so that libraries can mount their own collections with a minimum of technical support. Just as libraries formed consortia for

sharing cataloging information, institutions will band together to share their electronic texts, but the model will likely be a distributed one rather than the centralized OCLC and RILIN models for cataloging, with different institutions making available different texts accessible to one another over the Internet. Each institution will maintain a server for its texts, and each institution will maintain clients that access all the other servers. If we are lucky, work on the Z39.50 protocol will have advanced to the point where it can be used with these e-text clients and servers. In the meantime, gopher servers will provide an interim solution.

The greatest obstacle to the flowering of electronic text consortia will be copyright restrictions. As long as only out-of-copyright texts are shared, it is "simply a matter of programming" to realize the scenario described above. But out-of-copyright texts will not satisfy the needs of academic curricula. What we can look forward to in the not-too-distant future is a resolution of the legal obstacles to the distribution of texts and the development of models of agreement between publishers and electronic text centers as distributors. Publishers will need to realize that they *can* profit from the sale and sharing of electronic texts; distributors for their part will need to develop foolproof authentication mechanisms to reassure the publishers that their texts are still under control. As the articles in this issue show, the electronic text movement did not originate exclusively in libraries, although many librarians have been involved in its development. It is likely that this movement will provide an arena for cementing the alliance between libraries and computing centers, with the library contributing its considerable skills in collection management, bibliographic control, and user service to the relationship, to complement the necessary technological contributions of the computing center.

Whatever the future holds for electronic texts, it will be an exciting challenge for all participants, creators as well as users.

---

Katharina Klemperer is Assistant Director for Systems Development, Harvard University Library.



# Center for Electronic Texts in the Humanities

Marianne I. Gaunt

*The Center for Electronic Texts in the Humanities was established jointly by Princeton University and Rutgers University in 1991 to provide a national focus for the development, dissemination, and use of electronic texts in the humanities. The Center's primary activities include documenting existing electronic texts, developing a core collection of scholarly texts for access on the Internet, testing and promulgating standards for text encoding, and providing educational programs to support librarians, scholars, and teachers who are developing, maintaining, and providing access to electronic texts. The Center's activities have been supported by grants from the Andrew W. Mellon Foundation, the National Endowment for the Humanities, the New Jersey Committee for the Humanities, and the Booth Ferris Foundation.*

## History

In 1991 Princeton and Rutgers universities announced the establishment of the Center for Electronic Texts in the Humanities (CETH), a collaborative project between the state university and the major private university in New Jersey. CETH was to provide a national focus for those involved in the creation, dissemination, and use of electronic texts in the humanities.

The concept for the Center grew from nearly eight years of work spent at Rutgers University Libraries in developing the *Inventory of Machine-Readable Texts in the Humanities* and the needs that emerged from that experience. The *Inventory*, begun in 1983 with funding from the Council on Library Resources, provides an online catalog of electronic texts that can be of potential use to researchers in various humanities disciplines. It is maintained on RLIN, the online network of the Research Libraries Group, and is available internationally. By locating texts through the *Inventory* a researcher can save considerable time because the work of transcribing and encoding a printed text into an electronic format was, and still is, time-consuming and tedious, detracting from research time that could be spent using the text. The *Inventory* permits large-scale resources to be shared by many who have differing theoretical objectives and are working in diverse computing environments. With the exception of the Oxford Text Archive, begun in 1976, there have been few systematic efforts to make existing electronic texts available for other scholars to use.

Earlier compilations of texts were done in an ad hoc fashion for individual or group research projects or were

the by-products of a concordance, dictionary, or critical edition publication. There were no recognized procedures for providing access to the texts by others or maintaining them for the long term. There were few, if any, commercially published texts. As a result there were no published bibliographies to document the existence of the texts, as they were in the hands of a few individuals or research institutes. CETH estimates that approximately 95 percent of existing texts are in this uncataloged form.

When sources of texts were identified and compilers asked to send documentation from which a catalog record could be created, several other challenges arose. Most projects were not staffed adequately to provide the desired information. Both large and small projects did not foresee their data being used elsewhere and had not documented their texts extensively. In many cases, when the *Inventory* survey data were received, the information was incomplete or difficult to use. Reference to a source text upon which the electronic version was based was often missing and the encoding practices varied widely, usually depending upon the focus of the particular research project. It became increasingly clear that the reusability of these texts would depend heavily on the documentation of their quality and that standards for encoding and interchange were desperately needed.

After gathering data for the *Inventory* and after discussions with humanities scholars, other barriers toward advancing scholarship in the humanities through the use of high-quality electronic texts emerged: difficulty in locating electronic texts that could be used or adapted for use in research and/or teaching; a lack of standards for text encoding and interchange that would allow high-quality texts to be produced and shared by others who might or might not be using the same software; the need for better software tools that would advance methodologies in humanities computing beyond those in use for the last thirty or so years; and the lack of educational programs for those interested in developing skills in using computing for humanities research.

## Planning Conference

A number of professional associations were discussing the same problems and working on various aspects, such as text-encoding standards, but it seemed that a single focused operation might further the development of

**Marianne I. Gaunt** is Associate University Librarian, Research and Undergraduate Services, Rutgers University Libraries, New Brunswick, New Jersey. Her e-mail address is [gaunt@zodiac](mailto:gaunt@zodiac).

## What Is the TEI?

The Text Encoding Initiative (TEI) is a cooperative undertaking to define an encoding and interchange format for electronic texts. It is sponsored jointly by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). Due to the academic nature of the sponsoring organizations, the encoding scheme has been developed primarily with academic or scholarly uses in mind. The TEI standard provides a markup scheme to support literary, linguistic, historical, statistical, and other kinds of textual research.

Work on the standard was begun in 1987, when an international group of thirty-one scholars assembled at Vassar College. Out of this meeting emerged the TEI project. The project began with a two-year cycle that resulted in the first draft of its standard (affectionately called "P1"), which was published in June 1990. The second cycle has consisted of a thorough review and refinement of the initial guidelines. Most of the parts of the new guidelines have been made available via servers on the Internet. The complete "P2" version of the TEI guidelines should be appearing in printed form very shortly.

Very early in the project, the TEI identified SGML (Standard Generalized Markup Language, ISO 8879) as the vehicle for its work. The TEI guidelines should be thought of as a specific application of SGML and hence as conforming with the broader standard. For the TEI, SGML offers a number of advantages.

### Descriptive Markup

Most kinds of markup systems are procedural, specifying how a section of text should be rendered or presented, either on screen or on paper. By contrast, SGML-based markup is descriptive, simply identifying the boundaries of structures within a text. The advantage is that TEI-conformant texts can be processed by different applications and text-processing systems without the loss of encoding information. In addition, descriptive markup tends to focus on the logical structure

of a text, which, in most cases, is much more valuable to the scholar than a specific kind of typographical rendering. Procedural markup is fine for printing, but is far less helpful for pattern searching or stylistic analysis employing statistical programs.

### Document Types

SGML supports the concept of document types. Every SGML document has a document type declaration (DTD), which is a formal definition of the tags used in that document and their structure. The advantage here is that applications and programs can be written that can process documents of a certain type in the same way. Applications can also "specialize," providing unique processing capabilities for documents of a specific type.

The entire work of the TEI is contained in its DTD files. These files define all the tags and entity references that can be used in texts conformant with the TEI guidelines. The DTD files also lay out models for text structure.

For example, prose works are assumed to have front matter (title pages, publication information, etc.), the main body of text, and back matter. Similar DTD files exist for many kinds of texts, including (among others) drama, verse, dictionaries, transcriptions of spoken texts, text-critical apparatus, and language corpora. Tag sets have even been defined to deal with the difficult problem of graphs, digraphs, and tree-structure diagrams.

### Extensibility

SGML is extensible, meaning you can define and use your own tags. Once your document has declared its use of a TEI DTD, you can define additional tags and entity references. This is an important asset for academic work, since different encoding schemes are needed to represent diverse research interests.

### Device Independence for Document Interchange

Computing systems have long been afflicted by the Tower of Babel syndrome,

using different conventions for representing characters and document structures. The successful interchange of electronic documents between different computers can still be a challenging and frustrating task. The TEI's use of SGML ensures that documents that conform to its standard can be successfully exchanged with no loss of information. In addition, conformant documents can be shared and processed by a group of scholars with no additional preparation.

An important aspect of interchangeability is that one scholar can build on another's work. Tagging texts can be a laborious and time-consuming exercise. It's a wheel that definitely should not be reinvented. A scholar, receiving a TEI-conformant version of Goethe's *Faust* can immediately add tag "threads" to the document without needing to reconstitute the basic tagging already present.

### Recognized Standard

SGML was already an international standard by the time the TEI was begun. SGML offered the advantages of an established system, such as an existing base of computer applications and a wealth of literature explaining the way that SGML works.

The project has been generously supported by the National Endowment for the Humanities, the Directorate XIII of the Commission of the European Communities, and the Mellon Foundation.

There is a ListServ discussion group devoted to the TEI. To get a complete index of all the available files, send a message to [listserv@uicvm.uic.edu](mailto:listserv@uicvm.uic.edu). The text of this message should be a single line reading *index TEI-L*. This index will supply information about accessing an introductory "package" of files. This package contains information on the status of the project, how to access the additional online resources, and how to contact the editors of the TEI guidelines and the sponsoring organizations.—*Malcolm B. Brown, Director of Academic Computing, Dartmouth College, [malcolm.brown@dartmouth.edu](mailto:malcolm.brown@dartmouth.edu).*

effective methodologies for identifying, compiling, maintaining, and using electronic texts. A colleague at Princeton University who had contributed information to the *Inventory*, Robert Hollander, professor of Italian language and literature and director of the Dartmouth Dante Project, also recognized the need for such a focused effort. Together the director of the *Inventory* and Hollander organized discussions involving both universities on developing the concept of a national center that would address the issues that limited researchers from fully exploiting the potential of computing for humanities scholarship.

In March 1990 a planning conference for fifty international participants was organized, which was funded by NEH and the Andrew W. Mellon Foundation. The purpose was to test the concept of a national center for machine-readable texts in the humanities. A concept paper was sent to attendees in advance, which proposed the Center and described its activities and organization. The paper formed the basis for discussion at the two-day meeting.

The concept was greeted with enthusiasm, but there was no consensus on the activities or organization of the proposed Center. The list of potential activities was staggering because the needs were so great. There were, however, two major points of agreement: documenting the existence of electronic texts was critical, and providing educational programs in a variety of related areas was important, for more often than not, humanities-computing scholars may have no other colleagues of similar interest or experience on their campus. There was considerable support for the idea of operating the Center from a library, as participants realized both the institutional stability of such an base of operations and the natural links with other information resources.

There was much less agreement on the concept of an archive, which was one of the original activities proposed for the Center. The Oxford Text Archive was still willing to accept, preserve, and make available under compiler agreements any of the texts sent to it. Duplicating this effort did not seem necessary.

Following the conference the staff at Rutgers and Princeton synthesized the issues raised and deliberated on an appropriate model to meet each institution's own needs as well as the perceived national needs. The enthusiasm and interest generated during the conference were encouraging, but there was a need to evaluate all the comments in light of what could be achieved given likely funding and staffing. From the beginning, CETH was envisioned as a national node on an international network of projects and centers actively involved in the creation and dissemination of electronic texts to support humanities research and scholarship. While Princeton and Rutgers would be funding the operation for the first few years, the concept of a consortium of those institutions interested in the Center's work would be essential for its continued operation. A

consortium could provide an ongoing funding base and potentially serve as a distributed network for CETH operations and mutually agreed-upon projects. The unique contributions of CETH would come from its focus on the specific areas that would enhance humanities scholarship through the use of computing methodologies. CETH would provide an effective framework for the uses and users of electronic texts in the humanities.

In August 1990 Rutgers and Princeton submitted a joint proposal to the National Endowment for the Humanities to establish the Center. The request was not funded as submitted. NEH agreed to fund a full-time cataloger to continue the work of the *Inventory* and passed along comments from reviewers, which were helpful in framing a subsequent proposal. NEH recommended hiring a director who would commence operations to demonstrate how the Center would operate. In September 1991 that happened, and in July 1993 the Center for Electronic Texts in the Humanities was funded by a three-year grant from the NEH to carry out the activities described below.

## □ CETH's Activities

There have been some changes in the thinking about the activities proposed for CETH from those proposed at the planning conference only a few years ago. The changes come from a recognition of what is achievable as well as subsequent developments in humanities computing. CETH has focused its activities on three major areas: documenting electronic texts through the continued development of the *Inventory of Machine-Readable Texts in the Humanities*; developing a focused collection of high-quality texts encoded in standard format to be made accessible on the Internet and searchable with software developed for humanities-computing applications; and providing educational programs that discuss issues of software applications, standards for text encoding and documentation, and computing methodologies. While these may appear to be separate and discrete activities, they are truly interdependent. The framework that CETH expects to provide to equip scholars with much-improved methods to access electronic texts and thus enhance scholarly inquiry is dependent on the simultaneous accomplishment of all three goals.

## □ The *Inventory*

A separate article in this issue discusses the *Inventory*. This is one activity that has remained constant over

## What is SGML?

As its full name suggests, SGML (Standard Generalized Markup Language) is a system for the descriptive markup of electronic texts. To explain briefly what SGML is, the terms in the name are defined:

**Standard:** SGML is recognized by the International Standards Organization (ISO) and bears the ISO number 8879. While some may argue just how rigorously the SGML standard is observed, it is standard enough to be supported by a number of computer programs and to be adopted by a growing number of commercial and government organizations.

**Generalized:** This means that the kind of markup that SGML uses is not idiosyncratic to a word processor, a vendor's brand of computer, or even an operating system. This kind of markup identifies the functional components or the logical structure of a text, rather than determining how a text should be presented or typeset. It does so in a way that is sufficiently generic so that texts marked up using SGML can be exchanged reliably between computer systems, irrespective of word processor and operating system.

Anyone who has moved text files from one kind of computer to another (such as from a mainframe to a personal computer, or even from a PC to a Macintosh) knows the unpleasant surprises that can await one at the final destination: characters change. In addition, it is often not possible to preserve the formatting information (centering, emphasis, footnotes, etc.) when you move texts from one system to another. Word processing vendors have established interchange formats (such as Microsoft's Rich Text Format), but these systems are generally limited to that vendor's software products. SGML solves this dilemma by employing its generalized (or "generic") markup system.

**Markup:** The "markup" of a text is a way of making its structure as explicit as the words that make up the text. Numerous markup systems have been devised, many predating computing systems. What they have in common is that they all add codes to the words in a text in order to identify the text's structures or to determine the format of a text. Most markup systems concern themselves with the presentation of a text, specifying how the text should look when printed. By con-

trast, SGML's markup focuses on structure.

**Language:** SGML may be thought of as a kind of programming language for texts. It requires that all tags and entities in a document be formally declared and defined. These definitions give rules that regulate the use of the tags, and to be conformant, a text must observe these rules. SGML conformance is checked by a program called a parser. The "approval" that is given a text by an SGML parser is similar in many respects to the approval given to program code by a compiler.

But what does SGML look like? SGML uses a system of flags or "tags." The basic idea of tagging is very simple: an initial tag identifies where a structure (such as an act in a play, a stanza in a poem, or an emphasized expression) begins and a corresponding closing tag marks the end of that structure. For example, paragraphs might be "tagged" using the markers "p" and "/p," and emphasis might be identified using "emp" and "/emp." Tagged structures can be nested within other structures, as long as no overlap occurs.

The use of the term "emphasis" instead of "italics" may appear to be splitting hairs, but in fact points at a very fundamental distinction. Most kinds of markup systems are *procedural*, specifying how a section of text should be rendered or presented, either on screen or on paper. By contrast, SGML-based markup is *descriptive*, simply identifying the boundaries of structures within a text. Since there are several typographical conventions for typesetting emphasized words, using the generic concept of emphasis identifies the structure, leaving the question of how to represent it open. Additionally, the expression "might be tagged" is also deliberate. Although most uses of SGML supply a standard tag nomenclature for common text structures, these tag "vocabularies" are extensible. You are free to define your own tags, as long as you observe SGML's basic rules.

SGML also solves the difficult problem of computer character sets. While most computing systems (although not all!) agree on how to represent the basic alphanumeric characters, anarchy reigns when it comes to vowels with diacritics, em dashes, and so forth. This problem is even worse when it is necessary to repre-

sent characters that do not occur in the Latin alphabet, such as those in Greek, Cyrillic, and Asian alphabets. To solve this problem, SGML defines a system of what are called *entity references*, an unambiguous way of identifying a specific character. In SGML, an em dash might appear as "&emdash;" and an "o" with an umlaut as "&ouml;."

It may be objected here that to inject into a text unsightly beings like "/emp" and like "&ouml;" renders a text illegible to the human reader. It is true that texts marked up in SGML take on a degree of homeliness. But programs exist that can hide the SGML markup from the reader, or will interpret it typographically when the text is presented to the reader, either on the screen or printed on paper. The markup is done primarily for the "edification" of the computer. Its presence in the text is essential, since it provides the explicit codes a computer must have to do something interesting with a text: typeset it, search for instances of two terms coinciding within the same paragraph, or include an image that is not part of the text file itself.

Perhaps the main thing to remember about SGML is that its markup identifies the structures that make up a text and not its appearance. This logical and unambiguous markup of a text provides the basis for performing a number of computer-based functions, and hence provides the actual leverage we need when we use electronic texts.

For additional reading on SGML, see:

Bryan, Martin. *SGML: An Author's Guide to the Standard Generalized Markup Language*. Addison-Wesley, 1988. A detailed account with heavy emphasis on publishing.

Goldfarb, Charles. *The SGML Handbook*. Oxford Univ. Pr., 1991. An exhaustive presentation of all aspects written by the father of SGML.

van Herwijnen, Eric. *Practical SGML*. Kluwer, 1990. Good introductory textbook with emphasis on how SGML is currently being used.

Malcolm B. Brown. *Director of Academic Computing, Dartmouth College; malcolm.brown@dartmouth.edu.*

many years and is still considered the backbone of CETH. Since 1983, when the *Inventory* began, a number of commercially available texts have appeared. Many libraries are beginning to acquire and catalog them, CETH included. While this number will undoubtedly grow, so too will the number of potentially valuable texts compiled by individuals and groups for major research projects. Their existence and availability also need to be documented in the *Inventory*. While cataloging these texts may be accomplished in a distributed environment with many libraries sharing the work, CETH remains committed to gathering information about the existence of these texts on an international basis. In the future CETH hopes to link scholars directly from the *Inventory* to the texts they need.

## Scholarly Text Collection

Plans are under way to begin the second activity of the Center, which is to provide Internet access to high-quality texts with retrieval software designed for humanities applications. CETH does not propose to act as an archive or a repository for all humanities material in electronic form, but intends to make available a focused collection of scholarly texts maintained in the Text Encoding Initiative's (TEI) implementation of the Standard Generalized Markup Language (SGML). Users should expect the same quality of material they find in a printed source with adequate tools for accessing it.

The development of standards for the encoding and interchange of humanities texts was being addressed almost at the same time that the concept of a Center for Electronic Texts in the Humanities was being introduced. The Text Encoding Initiative (TEI), a major international project sponsored jointly by the Association for Computers and the Humanities, the Association for Computational Linguistics and the Association for Literary and Linguistic Computing, has promulgated standards for text encoding and interchange and, most important, for text documentation in the header of the electronic text. Commercially published texts are emerging, usually with good documentation, and a few publishers (notably Chadwyck-Healey) have started using TEI guidelines. However, many researchers still do not know about the TEI or how to implement the guidelines. Beginning with texts from the Women Writers Project (WWP) at Brown University, CETH will provide access via the Internet to focused collections of texts encoded in the TEI format. The WWP texts provide a perfect beginning for the text collection because they are in English, they are scholarly in nature, they will appeal to numerous disciplines, and they are encoded in TEI format. A large body of high-quality data on the network with

substantial user interaction is necessary to observe and test how the data is used. This information will form the basis for future research undertaken at CETH to improve software applications. Robin Cover, a member of the TEI Text Representation Committee, has been commissioned by CETH to prepare the initial specification of the software for CETH's text collection and to evaluate existing programs. CETH will hire a text systems manager shortly and work will begin in earnest on the development of this phase of CETH's activities.

At the same time that CETH is working on making these texts available on the Internet, the staff is collaborating with the School of Communication, Information and Library Studies at Rutgers on a research project to examine the uses and users of electronic texts. By using the CETH text collection, the project will develop models for future software design that will advance humanities scholarship by addressing how humanists interact with text. Humanities-computing methodologies have remained constant over the last thirty years, with researchers often limiting research queries to those that could be addressed by the software available. The new text-encoding standards combined with more sophisticated software have tremendous potential for humanities scholarship. In order for electronic texts to truly move into the mainstream of humanities scholarship we need to develop software that will address questions that humanists want to ask.

## The Summer Seminar

The third activity of the Center is the development of educational programs to provide support for librarians, scholars, and teachers who are working with electronic texts; to provide an opportunity to promote standards for text encoding and documentation; and to demonstrate the value and use of the TEI guidelines. For the last two years CETH has been offering a summer seminar called "Electronic Texts in the Humanities: Methods and Tools." The intensive two-week seminar addresses a wide range of challenges and opportunities that electronic texts and software offer. The thirty participants are chosen on the basis of applications describing their current work and their ability to act as resource persons at their home institution. The seminar covers such areas as data capture, markup, retrieval, presentation, transformation, and analysis of electronic text. Hands-on experience with various software is included, which is why the seminar has a limited enrollment. Everyone attending is encouraged to bring a sample of the project on which he or she is currently working. The seminar has been cosponsored by the Centre for Computing in the Humanities at the University of Toronto and is

held each year at Princeton University. Willard McCarty, a professor at Toronto, and Susan Hockey, director of CETH, are the primary instructors. Scholars involved in major projects also give lectures, and participants have an opportunity to hear how "real" projects are organized and operated.

## Support Activities

In the course of gathering information for the *Inventory* the staff has gained access to considerable amounts of information regarding the status of national and international projects and activities in humanities computing. As a result, CETH is in the unique position to act as a clearinghouse for information on projects, texts, and software related to humanities computing. The *CETH Newsletter*, produced twice per year, describes the activities of CETH and highlights other activities of significance to CETH's operations. In 1992 CETH established an electronic listserv for those interested in CETH's activities or for individuals to communicate directly with CETH staff.

To subscribe, send an electronic-mail message to Annelies Hoogcarspel, operator of the list, at: hoogcarspel@zodiac (Bitnet) or hoogcarspel@zodiac.rutgers.edu (Internet) with the one-line message only: SUBSCRIBE CETH. Hoogcarspel is also the moderator of a recently established distribution list for the ACRL Discussion Group on Electronic Text Centers called ETEXTCTR. This discussion group was a direct outcome of the first summer seminar at which the librarian participants wanted to extend their experience, share what they had learned with other librarians, and discuss issues related to the development of electronic text centers in libraries. The recommendation for the distribution list emerged from the first meeting of this group at the American Library Association Annual Conference in June 1993.

## Staffing and Organization

The work of the Center is currently carried out by three full-time staff members: Susan Hockey, director; Annelies Hoogcarspel, cataloger; and Christine Bohlen, library associate. While the main administrative offices of the Center are in the Alexander Library at Rutgers, the director has an additional office in the Firestone Library at Princeton and divides time between the two universities. There are full-text workstations at both library sites with a variety of materials for users. Three graduate students work part-time at the Alexander Library maintaining the workstations, preparing user aids, and assisting in training

librarians and patrons in using the materials. Each student has particular foreign-language capability. A text manager is currently being recruited and will become the fourth full-time staff member. The activities of CETH are shared between Princeton and Rutgers with each institution providing human resources, computing support, cataloging support and infrastructure, and space to achieve CETH's goals.

The Center is guided by a governing board and an advisory board. The governing board is composed of members from the library, the computing center, and two humanities departments at both institutions. It meets about three times per year, alternating between Princeton and Rutgers, with the goal of ensuring that the Center meets local needs. The advisory board meets once per year and is composed of fifteen members representing the international communities of humanities computing, information technology, publishing, and library science. The advisory board members provide broad advice to the Center staff on future directions, including collaboration with other national and international efforts that would further CETH's goals.

CETH is funded by Princeton and Rutgers with additional major grants from the Andrew W. Mellon Foundation and the National Endowment for the Humanities. Initial plans for CETH called for the development of a consortium after the first few years of operation. The purpose of the consortium is to provide an opportunity for institutions with similar interests to see the work of CETH moved forward more quickly, to provide a framework for project collaboration, and to provide a supplemental long-term funding base. There may be several levels of consorcial participation and CETH is currently working with the advisory board to determine the most suitable model.

## Conclusion

During the first year of operation, the director focused attention on personally networking with various research institutes, organizations, and individuals to raise awareness of the issues associated with developing scholarly applications in humanities computing and CETH's goals for addressing them. Where there is mutual benefit the Center will forge links with the many institutions and projects that already exist in the humanities-computing arena here and abroad and we hope that the phased development of a consortium will strengthen these links. CETH differs from many of these projects and organizations in that its primary objective is to act as a catalyst for the development of effective methodologies and procedures for handling and using electronic texts. Each of the activities in which CETH is engaged is focused on achieving

these goals. By situating itself in the library environment, unlike most of the existing major research projects and archives, CETH will capitalize both on library expertise in managing information and the humanists' expertise in developing and using it.

## Bibliography

- Hockey, S. "Developing Access to Electronic Texts in the Humanities." In *The Evolving Virtual Library: Visions and Case Studies*. Ed. L. Saunders. Westport, Conn.: Meckler, forthcoming.
- Hockey, S., and D. Walker. "Developing Effective Resources for Research on Texts: Collecting Texts, Tagging Texts, Cataloging Texts, Using Texts and Putting Texts in Context." *Literary and Linguistic Computing* 8, no. 4 (1993).
- Hoogcarspel, A. "Bibliographic Control of Electronic Texts." Paper presented at ACH-ALLC93 Conference, Georgetown, June 1993.
- Horowitz, L. "CETH as a Field Experience." *CETH Newsletter* 1, no. 2 (1993): 8-10.
- Sperberg-McQueen, C. M., and L. Burnard, eds. "Guidelines for the Encoding and Interchange of Machine-Readable Texts." Draft version 1.1, TEI document P1. Chicago and Oxford: ACH-ACL-ALLC, 1990.

# Collection Management for the 90's and beyond

**BLACKWELL'S**  
**NT|Online**

Blackwell's **New Titles Online** database includes **Tables of Contents** and **Descriptive Summaries** for new and forthcoming scholarly monographs.

- Browse the contents pages before placing your order
- Examine the publishers title description
- Order the title electronically
- Insure title coverage on approval
- Search both Blackwell North America and B.H. Blackwell New Title databases
- View full MARC Records
- Access through the Internet

FOR A FREE TRIAL OF THE WORLD'S MOST COMPREHENSIVE  
ACADEMIC COLLECTION MANAGEMENT SYSTEM,  
CONTACT YOUR BLACKWELL REPRESENTATIVE.

**BLACKWELL**  
NORTH AMERICA, INC.

6024 SW Jean Road, Building G, Lake Oswego, Oregon 97035.

Telephone: 800-547-6426 Fax: 503-639-2481

E-mail address: [BRIDGES@BNAME.BLACKWELL.COM](mailto:BRIDGES@BNAME.BLACKWELL.COM)

---

TECHNOLOGY DEDICATED TO THE LIBRARIES OF TOMORROW



# "A Library and Apparatus of Every Kind": The Electronic Text Center at the University of Virginia

David M. Seaman

*The Electronic Text Center at the University of Virginia combines an online archive of thousands of SGML-encoded electronic texts, all available through a single piece of search software, with a library-based center housing hardware and software suitable for the creation and analysis of text. Through ongoing training sessions and support of individual teaching and research projects, the Center is building a diverse and expanding user community locally, and providing a potential model for similar enterprises at other institutions.*

In August 1992, after almost a year of planning, the University of Virginia (UVA) Library opened its Electronic Text Center and online collection of full-text databases. Since then it has established itself locally as a catalyst for the development of an increasingly sophisticated user community, and nationally as a model for the large-scale implementation of online texts in higher education.

## Underlying Principles

Virginia's electronic text enterprise is founded on several underlying assumptions and expectations, which the first eighteen months of operation have tested and strengthened. Principal among these are the notions that—whenever technically and legally possible—the electronic texts should be remotely accessible, encoded in a uniform, nonproprietary manner, and all online data should be delivered up through a single piece of search-and-display software. As have other libraries, the University of Virginia has seen the negative side of a growing collection of e-texts on CD-ROMs: they are difficult to network on a campus-wide scale, they are discrete items that cannot be intermingled in a larger data set, and they typically come with their own search tools. The prospect of having to teach and support dozens of search packages, all doing much the same thing, was not a welcome thought. By buying the data on disk or tape and loading it into an online system that is under local control (and can therefore be tailored somewhat to local needs), the Center has largely avoided the logistical nightmare of working with electronic "books," each of which has its own access methods.

Moreover, it was an assumption from the beginning that simply providing e-texts and related technologies was not enough: they had to be offered up in a manner appropriate to users whose needs are often sophisticated but

whose technical experience is often limited, as is their sense of the possibilities of computer-aided textual study. Therefore, these new pedagogical and research services had to be introduced and nurtured through ongoing training sessions, orientations, demonstrations, and online help. The Center actively recruits new student and faculty users when there are holdings in their areas of study, and introductory sessions are tailored to a particular course or seminar, focusing on, say, just the eighteenth- and nineteenth-century philosophy holdings or just the Middle English poetry and prose.

## The Center

The Electronic Text Center is open and staffed most of the time that the library is open and provides a place where users can be introduced to the new services, where existing users can get regular help, where offline electronic texts are available, and where scanners and text-analysis software can be used. Having a walk-in center is central to the ability to provide electronic text technologies to university patrons: it is certainly possible to have an online text archive without a supporting center, or a library-based center with no online component, but the impact of these two items together not only has proven to be much greater than either one in isolation, but may be the only practical way to create a broad-based user community that sustains the growth of the service.

The Electronic Text Center, therefore, is both a work space for the coordinator and a staff of graduate students, and a place for users to learn how to create e-texts, to operate a range of related text-analysis software, and to access those e-texts not available online, including the following: the works of Immanuel Kant; the Global Jewish Database; the ICAME language corpora; Perseus (a collection of Greek texts and images); CETEDOC, a database of Latin theological works; and a hypertext edition of the works of Thomas Aquinas. The Center is equipped with three MS-DOS machines, a NeXT machine, a Macintosh, an IBM RS/6000, two additional X-Windows terminals, scanners that turn printed text into computer-readable forms or produce digitized images, laser printers, CD-ROM drives, and software that can generate indexes,

**David M. Seaman** is Coordinator of Electronic Texts, University of Virginia Library.

The title quotation comes from a letter from Thomas Jefferson to Henry Lee, Monticello, dated May 8, 1825: "The last donation of the legislature to the University was appropriated specifically to a library and apparatus of every kind." Thomas Jefferson, *Writings* (New York: Literary Classics of the United States, 1984), 1,500.

collations, concordances, word lists, statistical analyses, and hypertexts. Image-viewing software allows one to work with digital images of manuscripts, typescripts, book illustrations, and related visual materials alongside the searchable databases (images that may well have been created by the user in the Center).

## The Online Archive

The online texts are stored on an IBM RS/6000 workstation that currently contains the following full-text items: *Oxford English Dictionary*, second edition; the entire corpus of Old English writings; selected Library of America titles; several versions of Shakespeare's works; hundreds of literary, social, historical, and philosophical materials in several languages (chiefly from the Oxford and Cambridge Text Archives and from Oxford University Press); the British philosophy texts and other items from IntelLex Corp.; and the currently released parts of two massive databases: J-P. Migne's *Patrologia Latina* and the English Poetry Full-Text Database (when complete, this last item will contain every poem by every poet in the *Cambridge Bibliography of English Literature*). These texts are available twenty-four hours a day to users and can be accessed from the campus, from home, or from wherever else the user happens to be. Because of contractual obligations with the vendors who supply the texts and the search software, access to the online text service is not available to the whole Internet; it is limited to University of Virginia students, faculty, and staff. Understandably, the commercial purveyors of e-texts and search software want to sell copies to other institutions, too.

## SGML Markup

All of the online texts are encoded with Standard Generalized Markup Language (SGML).<sup>1</sup> The commercial databases usually come already tagged, and the others are marked up at the Center with the aid of volunteers from various library departments, following the implementation of SGML being formulated by the Text Encoding Initiative (TEI). The public-domain texts also often require time-consuming checking for completeness and accuracy and an amount of bibliographical sleuthing to determine what edition has been used to create the electronic version. Too rarely is this information already attached to the text, and the first thing that must be done is to affix a header to the document, recording its provenance, its print source, and the work that has been done to it locally. It is difficult for anyone to make appropriate classroom or research use

of an electronic text until he or she knows if the electronic record is a reliable (although perhaps elderly) edition or a textual nightmare, a state-of-the-art production or a product of outdated textual theory. Over time, the Center will be making available to others the public domain texts that have either been created or checked at the Center.

The practical implications of SGML texts are several fold. First, the tags operate both as a series of database fields—descriptive markers—that are invaluable when constructing and limiting searches, and also as a set of typographic and layout instructions for the appearance of the text on the screen. Figures 1 and 2 show how in practice the SGML tags work in the databases. In figure 1, one can see the *Oxford English Dictionary* being searched for any word beginning with "technolog," that result being limited to the earliest-quoted usage field ("First Quot"), and a twentieth-century item called up, and then limited again to those times that "technolog" appears as an earliest-quoted usage in seventeenth-century quotations. Searches can be limited in this manner because the text is marked up heavily with SGML tags denoting classes of information (date, quotations, etc). Next to the window displaying the 1615 use of "technologie" is a view of the text with its SGML tagging visible. You can clearly see such tags as `<EQ>` and `</EQ>` bounding earliest-quoted usage, `<D>` and `</D>` for dates, `<A>` and `</A>` for authors, etc. Normally, a user would not see these tags, but would see instead the typographical representation the *OED* uses for that item—dates are in bold, authors' names in full and small caps, and so on.

Figure 2 shows the Modern English Texts collection and the English Poetry Full-Text Database being searched for words beginning with "read" near words beginning with "pleasur," and this combination being limited to works published between 1750 and 1850. Such "proximity searches" are often effective ways of tracking a concept (reading for pleasure) through a body of literature. Results are displayed (clockwise from right) from Frances Brooke's *The History of Lady Julia Mandeville*, volume 1, page 15; Thomas Warton the Younger's "The Progress of Discontent"; and Jane Austen's *Northanger Abbey*, page 106. In the lower-right corner is a digital image of the title page of *Lady Julia Mandeville*. This work of Frances Brooke's is one of several texts created by UVa faculty—in this case by Patricia Spacks, the chair of the English Department—in order to use them in the classroom.

SGML offers additional practical advantages for an electronic text project such as this one:

- The tags are part of the text and not the proprietary addition of a particular piece of software, so the text and all its structure will outlive any SGML-compliant software currently used to search and display it.

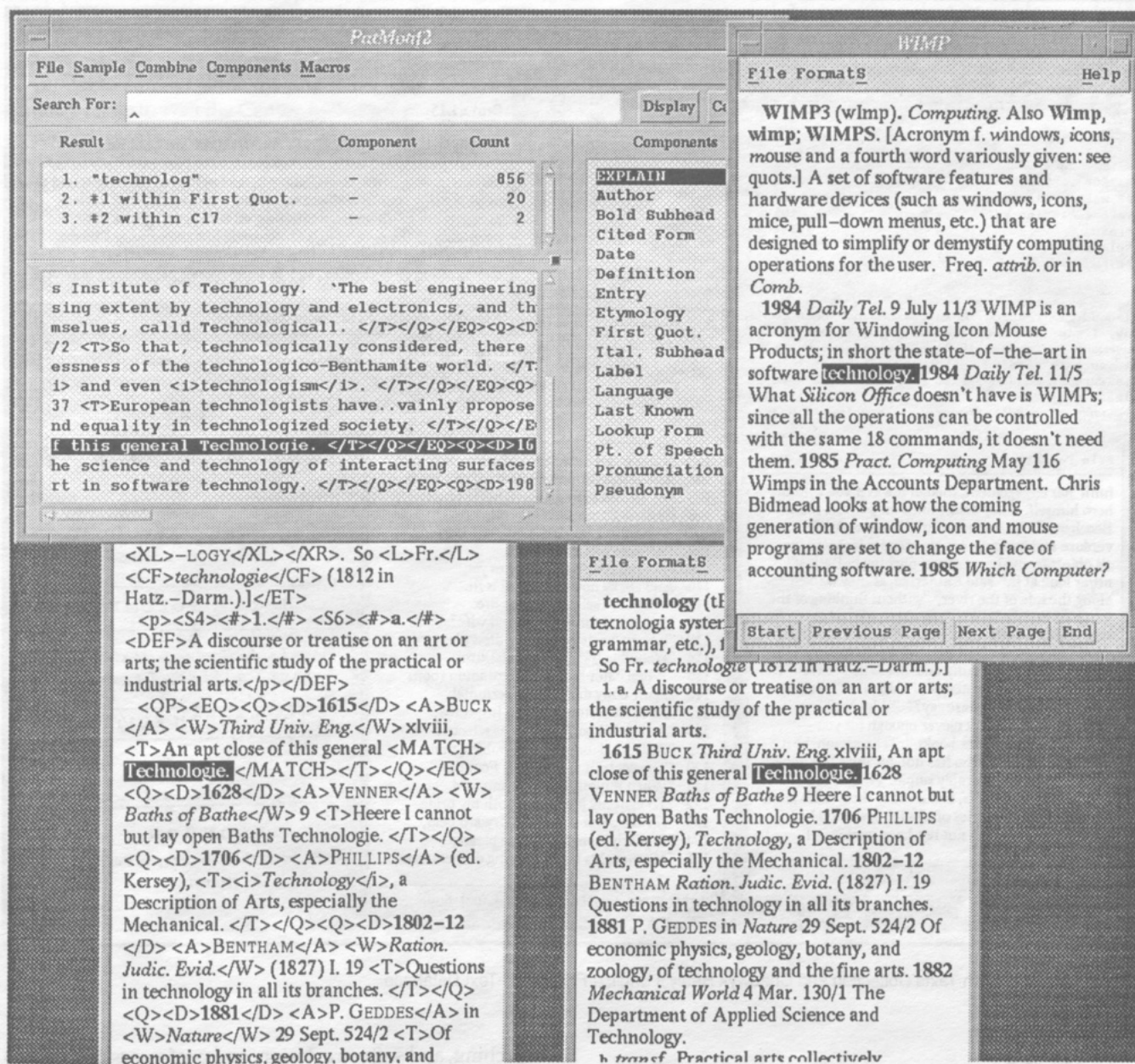
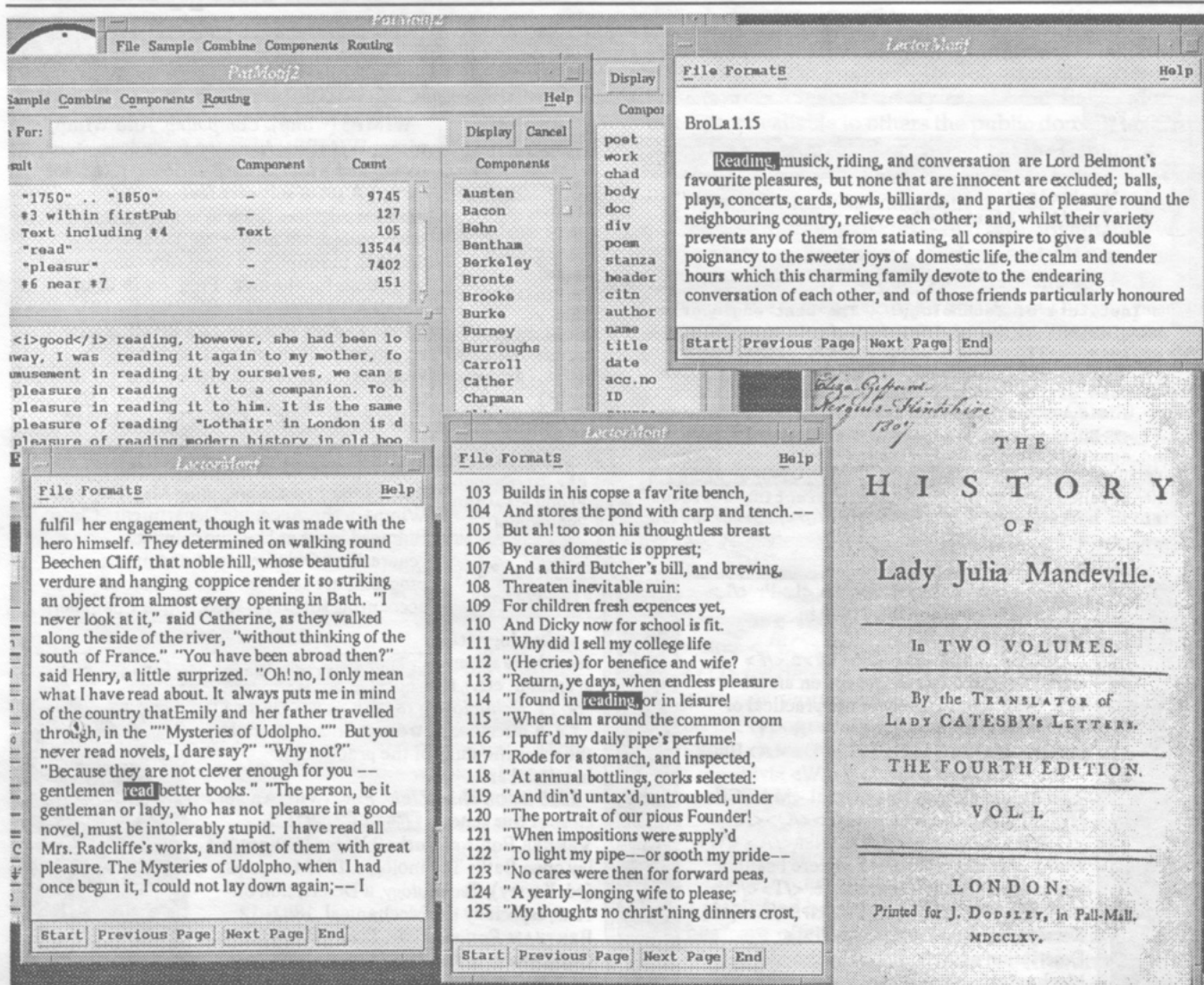


Figure 1  
Oxford English Dictionary

- A user under guidance can add a level of tagging specific to his or her particular needs (marking, say, particular grammatical or linguistic features).
- It is easy to add a new SGML text to an existing collec-

tion of texts (having the Riverside Shakespeare online is beneficial, but the text is much more valuable if the user can choose to search it alongside scores of other sixteenth- and seventeenth-century texts).



**Figure 2**  
 UVa's Modern English Texts Collection and Chadwyk Healy's English Poetry Full-Text Database

## Users

The first eighteen months of operation have seen a surprising number of patrons, ranging from first-year undergraduates in composition classes to graduate students studying aspects of Anglo-Saxon and Middle English literature, Shakespeare, analytical and descriptive bibliography, and the eighteenth-century novel; religious studies scholars working with the Hebrew Bible and rabbinical responsa; French students learning medieval French; graduate classes from both the Computer Science and English departments examining data encoding, text

searching, and software design; and professors or students from the History, Commerce, and Education departments working on individual research projects. In several instances faculty have been able to include in their courses electronic versions of texts that have long been out of print, and therefore inaccessible for classroom use, either by creating the texts themselves through the E-text Center and adding them to the online holdings or by using texts from a collection such as the English Poetry Database that the library does not own in print.

In addition to work with patrons at UVa, the E-text Center and online archive have been able to provide a

model for other institutions as they plan similar endeavors. After the Center received some regional and national notice last year, scores of non-UVA librarians and scholars have made contact with us, and a number have subsequently come to visit the Center, including parties from the following institutions: Harvard, Indiana, Johns Hopkins, Iowa, Yale, the University of Nottingham, Virginia Commonwealth University, Virginia Tech, Emory, the University of Richmond, Oxford, UNC Chapel Hill, Kansas, Groningen, Leiden, Macquarie University (Sydney), the British Library, and Duke. This activity makes it possible to aid the development of electronic text services elsewhere. It is hoped too that close liaison with the academic computing departments at UVA will provide an example of how to create an information technology community by unifying the creative energy and expertise of technical and nontechnical departments.

## Conclusion

The University of Virginia Library has established a large collection of online electronic text databases, delivered up through a single search program, and a support structure designed to build the collection and to foster its use. The library has been particularly effective in integrating the new suite of services into its existing structure,

making full use of the expertise already in place in fields such as cataloging, acquisitions, and collection development, and applying these skills to the purchase and bibliographical control of electronic texts. Although it is still early in this endeavor, there has already been some success in creating a body of users who are excited about what e-text technologies offer them, and who are making increasingly sophisticated use of the facilities. Judging from these early users, and from the interest this project has generated from other institutions, it seems reasonable to assert that online text collections and library-based centers for textual creation and analysis are becoming an increasingly significant part of the pedagogical and research resources that a university provides.

For more information about the Electronic Text Center, send e-mail to [etext@virginia.edu](mailto:etext@virginia.edu) or write to the Electronic Text Center, Alderman Library, University of Virginia, Charlottesville, VA 22903.

## References and Notes

1. For an introduction to SGML, see Eric Van Herwijnen, *Practical SGML* (Dordrecht, Netherlands, and Boston: Kluwer Academic Publishers, 1990). This item is also available as an electronic text. For a copy of the Text Encoding Initiative guidelines, ftp to the anonymous ftp site at [sgml1.ex.ac.uk](ftp://sgml1.ex.ac.uk) and choose the directory /tei.

# Building the Digital Library: The University of Michigan's UMLibText Project

Beth Forrest Warner and David Barber

*Over the past decade, new data formats and new tools have emerged that hold the promise of greatly enhancing the resources available to scholars. The UMLibText initiative is one of several evolving projects designed to provide expanded and/or enhanced access to electronic information resources for the University of Michigan community. Following a review of the history of this textual-analysis project, its current status and continuing development are discussed in the context of the rapidly emerging electronic information environment on campus.*

The much-publicized information explosion and technology revolution hold the potential for greatly enhancing the resources available to scholars. New data formats and new tools are continually emerging, but the scholar's knowledge of and ability to utilize these tools effectively may lag behind the rapid rate of development. Universities, and in particular libraries, must continue to develop the capacity and expertise to promote and provide access to these new systems and resources. An example of one such effort is the UMLibText project under way at the University of Michigan.

## Project History

During the mid-1980s, the University of Michigan Graduate Library became involved in several initiatives to support primary information in electronic formats. One of the first initiatives involved providing access to electronic text services such as ARTFL (American and French Research on the Treasury of the French Language) at the University of Chicago. Use of the ARTFL service soon raised sufficient awareness and interest in electronic text services that a local implementation of a similar service was desired. By 1988, several more initiatives had developed to the extent that the library recognized the need for a data services librarian. The position filled that year involved traditional reference service, advisory and technical support functions for bibliographic databases, and collection development responsibilities, plus responsibilities for further development of initiatives involving primary data in electronic formats.

During 1986-89, as the library gained more experience with textual analysis, it became apparent that no one software package could accomplish everything for all users—each package facilitates specific types of analyses. In addition, staff gained a clearer understanding of the nature of the work being done by researchers, i.e., it is not signifi-

cantly different from work done with print resources, it is intensely private work, and it cannot be mediated by a librarian. It was also quickly apparent that the environment for text and software was far from ideal—standards were only beginning to evolve for the description of text, and the software for use with the standardized texts was a rare commodity. Out of these early experiences, the library eventually developed three goals for its electronic text initiatives:

- To provide a software and hardware platform adequate for serving a broad range of textual-analysis needs
- To facilitate remote multiuser access
- To begin a local collection of text in a standardized format suitable for a variety of textual-analysis applications such as authoring studies, thematic analysis, and linguistic analysis

To further inform development of the local project, the data services librarian surveyed Research Libraries Group (RLG) academic libraries during mid-1989 to determine what role libraries were playing in supporting textual analysis on campuses.<sup>1</sup> Survey results indicated that a majority of libraries were actively involved in providing support for machine-readable text files, but most were not yet collecting such files or providing the necessary tools to use these files locally. Thus, the primary involvement appeared to be in helping users locate or use files—little library acquisition of files was taking place.

Building on the information gained from the survey, the University of Michigan Library launched its electronic text effort in 1989 as a pilot project, thus linking the activities of collection development and departmental liaison with data-support functions. The PAT software package was acquired and mounted on an early Sun SPARCstation lent to the library by an experimental computing unit of the university's Information Technology Division. PAT was originally developed at the University of Waterloo for use with the second edition of the *Oxford English Dictionary* and is currently marketed by Open Text Systems, Inc. It is a powerful search engine capable of handling large amounts of structured and unstructured text in a variety of fairly generic ways. PAT is a pattern recognition system and has both text-based and OSF/Motif-based user interfaces running under X-Windows. The Motif interface simplifies the query process somewhat and links into the LECTOR text display and formatting system when specific text selections are made.

---

**Beth Forrest Warner** is Head, Library Systems Office, and **David Barber** is Graduate Library Information Technology Coordinator, University of Michigan, Ann Arbor.

In an effort to acquire a critical mass of electronic texts quickly, the library negotiated the first institutional agreement with the Oxford Text Archive. This agreement provided several hundred texts as well as the approximately three thousand texts and text fragments of the Toronto Old English Corpus. A similar agreement was made with the Cambridge Text Archives, a much smaller repository of older German-language texts. The software and texts were pulled together, with assistance from the Library Systems Office, by using shell scripts and minimal C programming to create a fairly transparent system that did not require users to learn the UNIX operating system. System development was aided by the participation of a number of dedicated and interested users from various humanities disciplines who used the system, gave feedback, and generally helped the library understand the functions needed.

At the end of 1990, a proposal was put forward for an expanded umbrella project—the Center for Information Technology Experimentation (CITE). CITE was viewed as an opportunity to create a dynamic space in the library for highlighting new information technologies and providing instructional programs in the use of these tools. In addition to the electronic text efforts already under way, other initial CITE subprojects included easy access to external library catalogs, access to the UPI NewsWire service provided by Clarinet, a lexicographer's tool kit, electronic journal access, current U.S. Census Data, access to numeric data sets, and a geographic information system.

As part of the CITE effort, the UMLibText (UML) project, as it was now called, became a more formal initiative, using only library resources. The library leased the electronic version of the *OED* and purchased site licenses for PAT and LECTOR, allowing these packages to be made widely available to the campus. LECTOR is an OSF/Motif display package, also developed at the University of Waterloo and marketed by Open Text Systems, Inc. It allows users with X-Windows-equipped workstations to display search results as a high-resolution image closely approximating the original page, with formatting that includes boldface, italics, and variable-sized fonts. To support this system, the library purchased one of the first IBM RS/6000 Model 320s, three gigabytes (3 GB) of disk space, and one gray-scale X-terminal. The existing menuing system was enhanced, and users were able to access the system in either VT100 or X-Windows mode. The library further expanded the textual resources by adding several commercial texts and developed staff skills in text encoding. Continued publicity and presentation of the system as a production service soon brought UMLibText a high level of visibility on campus.

Early in 1992, through gifts by Digital Equipment Corporation and the university's College of Engineering, as well as through additional library support, the service

was expanded to include four more public X-terminals, a color DEC 5000, and physical facilities suitable for research and instruction.

Finally, as a benchmark for the services being provided by the University of Michigan Library, a follow-up survey to the 1989 RLG survey was done in 1991 to see what changes, if any, had occurred.<sup>2</sup> The results of this second effort revealed that the following had occurred during the intervening eighteen months:

- The number of libraries purchasing full-text for local textual-analysis efforts had more than doubled.
- Significant commercial offerings were being made available, as well as an increased number of commercial services (e.g., InteLex, Shakespeare on Disk, and Oxford University Press Electronic Publishing Division products).
- More analysis tools were being introduced (e.g., ETC's WordCruncher).

## Resources Offered

The holdings of the UMLibText system currently include the following text collections: the *Oxford English Dictionary*, second edition; the Old English Corpus; the first installments of the Chadwyck-Healey *Patrologia Latina* and pre-1900 English poetry text collections; a collection of Middle English works; a copy of translations of the Bible and of the Koran; the first folios and early quartos of Shakespeare; and a large collection of modern English texts that include both literary and philosophical works. The principal source of these texts has been a select number of electronic text vendors, including Chadwyck-Healey, the Oxford Text Archive, and InteLex. For the Oxford Text Archive materials, some SGML (Standard Generalized Markup Language) encoding was done by staff in the Graduate Library.

Texts can be searched as individual titles or in groups by author, date of composition, or intrinsic characteristics of the texts. Users can search for words, word stems or roots, single characters embedded in words, phrases, combinations of words, and conceptual word sets. A wide variety of display options are available, from single-line keyword-in-context listings to full-page emulation.

## Access and User Support

UMLibText is available to all faculty, staff, and students affiliated with the University of Michigan's Ann Arbor campus. To use the full resources of the system,

users obtain an account by submitting a request in writing, along with a signed acknowledgment of the copyrighted status of many of the texts. Valid requests are usually processed in less than one week. Selected text files are also available for use by anyone on the premises of the library. Five public X-terminals are provided for users in the Graduate Library. Personal accounts allow UML users to access the system either through a VT100 or an X-Windows interface. They also allow UML users to save or download portions of the texts. In addition, they permit the use of PAT/LECTOR with user-created text files. Having received an account, a UML user must then set up or configure his or her own equipment, with the assistance of library staff as necessary. This may involve modifying log-in files, or if X-Windows is to be used, getting the necessary fonts from the anonymous FTP site on the UML machine.

Users of the UML system are given a brochure describing UML when they receive their account name and password. That brochure points them toward the location of further user guides and bibliographies available through anonymous ftp. Guides currently available include information on using the VT100 and X-Windows interfaces, setup information for X-Windows, and bibliographies of the available texts. Occasional workshops are held to instruct system users and promote the UML system. Users who need assistance beyond the available guides and training sessions can send e-mail from within the UML system to library staff. UML texts and functions are regularly advertised as a part of library "technology fairs" on campus. Other special promotions are conducted when new text collections are acquired. A library guide describing UML is available in library-guide stands in the Graduate Library and other branches. Limited advertisement of, and access to, UML also occurs through a link from the library's gopher service, ULibrary.

User support for the UMLibText project is currently supplied by the Graduate Library Information Technology Coordinator while technical support is provided by an applications programmer in the Library Systems Office. In addition, joint development efforts are being pursued informally with staff at the University of Virginia.

## □ Use Data and Trends

The diversity of resources contained within UMLibText has ensured that the system is used by faculty, staff, and students from many different parts of the campus. The user community for UML includes not only literary and linguistic scholars, but also faculty, staff, and students from other humanities and social science departments, as well as from the Medical School and from Electrical Engineering and Computer Science. While this breadth of use may

be largely due to the availability of the *OED* through the UML system, it is also a result of a significant interest among the local information-technology community in structured text and electronic text systems.

The addition of a basic logging feature in the interface (X-Windows or VT100) allows the project to count sessions by user, date and time, and method of access. No attempt is made to measure the use of specific resources. The USERLOG files from UMLibText reveal that the VT100 interface is most frequently used. It also, however, reveals that the more frequent UML users tend to use the X-Windows interface. This pattern of use, and the additional fact that most UML users access the system fewer than ten times during a year, are at least partially accounted for by the existence of a few classes that teach about the UML system, including both Library School and literature classes. The most frequent users of UML are representative of the entire UML user community, not just the core literary researchers.

## □ Future Directions

### Availability of Encoded Text Collections

For the UMLibText initiative to grow, it is necessary to have a continuing supply of new textual resources available for local access and manipulation. While several major text collections are coming to market in raw form, others, such as the ARTFL project's French-literature resources, are made available only through specialized remote-access systems. These, together with the locally available texts, meet the needs of only a limited group of users.

In addition to the scarcity of texts, markup of those available to be loaded locally has become one of the biggest challenges encountered by project staff—the major linguistic corpora have been developed using a variety of text-encoding formats. For instance, most ancient Greek resources are issued in beta-encoded format, while other textual resources are appearing on the market using the SGML format. As more researchers have worked with electronic texts and become more familiar with the requirements and problems involved in computer-aided textual analysis, the need for use of a consistent encoding standard has become more apparent.

Recognizing this continuing limited supply of consistently encoded texts, two questions are raised. The first is, "Should we encode our own SGML texts locally?" While this may sound like a reasonable undertaking, staff quickly discover that SGML is a very complex formatting system and adequate tools to aid editing in SGML are expensive and therefore not easy to obtain—even if expertise in SGML is developed by a staff member.



---

The second question is, "What relationship will develop between the UML text collection and other text collections?" The electronic text information encoded using SGML is currently a small fraction of the total supply. In the meantime, many other significant text resources are becoming available online or as part of commercial software packages with texts in proprietary formats that may include warnings indicating that separating the text from the software constitutes a violation of copyright. If purchase of these types of electronic texts from publishers becomes common, the SGML-encoded text collection may become a very specialized, exceptional service. At that time, there may be pressure to move those texts to a more generic text system.

Inherent in these considerations is the underlying question of the responsibility of libraries to provide their communities with text that can be analyzed as variously and easily as the printed works available on our shelves. Libraries must commit more fully to building collections that conform with open standards—and encourage publishers to make more standardized encoded text readily available. The further development and use of SGML is seen as important for current as well as future use and compatibility of electronic texts, since it is an international standard and is not tied to a specific technology. In support of this direction, a major proposal outlining guidelines for text encoding has been formulated by the literary and linguistic computing community through the Text Encoding Initiative. In addition, similar guidelines for manuscripts have been published by the American Association of Publishers. While technological developments have made it increasingly possible to provide easy access to many simultaneous users outside as well as inside the library, our users may arrive only to find the promised texts unavailable, or unsuitable, for analysis.

### Client Software

While there is still need for a VT100 interface for remote access, client-server architecture is rapidly becoming the preferred model for provision of information services in most networked environments. A number of clients have been developed for the PAT/LECTOR software that underlies the UML system. Dartmouth College and Stanford University have developed Mac clients for PAT/LECTOR; and Open Text, the creators and marketers of that software, have developed a Windows client. Other clients are also available from other electronic text centers. Some of these clients have been developed for use with specific electronic text collections. As such, they provide interfaces that are good for specific products such as the *OED*, but not as useful for other collections. Some of the clients that have been developed are available free of charge, while others must be purchased. The challenge to the University of

Michigan Library is to find clients with the best interface for each of the principal operating systems.

It might seem as though we need only try out each of the various clients, searching for the best mix of Mac and Windows clients available. However, locating clients is not the only, or even main, problem. Just as in the beginning of widespread use of the Internet, when the number of network services proliferated along with the list of IP addresses you had to remember, the number of available clients is growing to the point where they are becoming a burden for information system users and providers. Faculty members or students might need to have on their desktop machine a UML client, a Mosaic client, a gopher client, a newsreader client, and a Z39.50 client—among others. Each of these clients might have its own user interface and configuration instructions and requirements. They may come from different ftp sources and have upgrades that occur on different schedules. One hesitates to contribute further to this complicated situation.

To avoid adding to the growing confusion, UML staff are currently considering promotion of a multipurpose client such as Mosaic. The Mosaic client is gaining acceptance on campus for access to gopher and World-Wide Web resources and has support for built-in forms that could be used to create a query to send to the PAT/LECTOR software. At least one PAT/LECTOR site has also created an SGML-to-HTML filter to translate the search output for Mosaic display (HTML stands for Hypertext Markup Language and is an SGML document type used for representing documents to be used with World-Wide Web servers and clients, such as Mosaic). The development of a true SGML viewer for use with Mosaic would further enhance its use. If Mosaic can be used as a client, service to the full-text collections could be expanded, since a central access method would exist in which users of other electronic services would also find out about UML resources. Instead of being a separate and partially hidden resource, UML becomes simply one more tool listed alongside other commonly available network tools.

### Authentication and Authorization

Access to the UMLibText system is currently available to anyone showing an official affiliation with the University of Michigan. To control that availability, the library has had to rely entirely on accounts administered by Library Systems Office staff on the local UML system server. Due to recent changes beginning to take place in the campus computing environment, the library should soon be able to take advantage of the unique authentication and Kerberos authorization services being widely promoted and implemented. This campuswide shift from a "one ID per system" to a "one ID per user" concept should greatly improve the ease of access for the University of Michigan

user community and, as a by-product, reduce the system administration overhead for the library.

## UMLibText in Context

As noted above, the UMLibText initiative is one of several evolving projects designed to provide expanded and enhanced access to electronic information resources for the University of Michigan community. As the number of resources has grown and the variety of access points and methods has proliferated, there has been a growing recognition of the need for a more coordinated approach. In November 1992, managers representing the University Library, School of Business Administration Information Resource unit, the Law School Library, the Information Technology Division, and the School of Information and Library Studies set a goal to identify how they might offer faculty, students, and staff coordinated information resources, systems, and support services relevant to their instructional, research, and administrative needs. A subcommittee was formed to draft a concept document to assist in discussion.

The final report of the subcommittee described a structure and process that eventually resulted in the formation of a joint initiative between the University Library, the Information Technology Division, and the School of Information and Library Studies to implement the digital library environment. This environment is envisioned as an organizing paradigm, a unified guide for the university community that suggests coherence, intellectual richness,

and purpose in the resources and services offered. If implemented well, the end result will be to empower its users by facilitating their access to, and retrieval of, information that meets their needs.

The underlying systems in this environment will be designed or selected to be standards based and highly flexible in order to adapt to the anticipated range and changing patterns of use. The Digital Library Project will present a unifying view of the University of Michigan's information resources while allowing individuals to use alternative access points to the information as they wish.

Internet navigation (using gopher, Archie, Veronica, or the World-Wide Web), bibliographic access (using the online catalog and citation files in MIRLYN or a table-of-contents service), statistical-data-file access (using Census and ICPSR files), geographic information services (such as ARCView), full-text retrieval, analysis, and display (through the TULIP electronic-journal project, the CIC electronic-journal archive, or UMLibText), plus many other services and resources—all are an integral part of the new, and evolving, information environment at the University of Michigan.

## References

1. John Price-Wilkin, "Text Files in RLG Academic Libraries: A Survey of Support and Activities," *Journal of Academic Librarianship* 17, no. 1 (Mar. 1991): 19-25.
2. John Price-Wilkin, "Text Files in Libraries: Present Foundations and Future Directions," *Library Hi Tech* 9, no. 3 (1991): 7-44.

GAYLORD  
INFORMATION  
SYSTEMS

**Gaylord**

The Trusted Source<sup>®</sup>

## Please send me more information:

- I want to know more about GALAXY<sup>®</sup>, Gaylord's fully integrated library automation system.
- I want to know more about Gaylord's SUPERCAT<sup>®</sup> CD-ROM Cataloging Module.
- I want to know more about Gaylord's Retrospective Conversion Services.

Name \_\_\_\_\_

Title \_\_\_\_\_

Institution \_\_\_\_\_

Address \_\_\_\_\_

City \_\_\_\_\_

State \_\_\_\_\_ Zip + 4 \_\_\_\_\_

Phone \_\_\_\_\_

Gaylord Information Systems

P.O. Box 4901, Syracuse NY 13221.4901

**1.800.962.9580**



NO POSTAGE  
NECESSARY  
IF MAILED  
IN THE  
UNITED STATES

**BUSINESS REPLY MAIL**

FIRST-CLASS MAIL PERMIT NO. 5869 SYRACUSE NY

POSTAGE WILL BE PAID BY ADDRESSEE

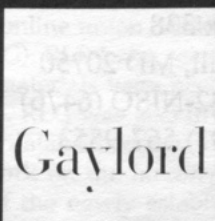
Sandy Schulman  
GAYLORD INFORMATION SYSTEMS  
Dept ITL  
PO BOX 4901  
SYRACUSE NY 13221-4901



# For the Latest In Library Automation Technology, There's only one thing you need to know:

Gaylord  
Information  
Systems

P.O. Box 4901  
Syracuse, NY  
13221.4901



The Trusted Source<sup>®</sup>

**Our phone  
number.  
1.800.962.9580**

Call and ask about GALAXY<sup>®</sup>,  
Gaylord's innovative, fully integrated library automation system.  
A system designed to meet the demands of any size library.

**Total Library Automation  
Made Easy.**



# NISO PRESS

Publisher for the National Information Standards Organization (NISO)

## **Z39.50-1992 Information Retrieval Service Definition and Protocol (Version 2)**

Z39.50 is at the heart of today's automated library systems. It makes it possible for any library to extend its reach beyond its own collection to tap the resources of remote collections and databases. Best of all, users don't have to learn the user interface of the remote system.

**Price: \$44.00**

**ISBN: 1-880124-07-6**

## **Z39.19-1993 Guidelines for the Construction, Format, and Management of Monolingual Thesauri**

This 1993 edition is the authoritative guide for constructing single-language thesauri, one of the most powerful tools for information retrieval. Written by experts, Z39.19 shows how to formulate descriptors, establish relationships among terms, and present the information in print and on a screen. Also included are thesaurus maintenance procedures and recommended features for thesaurus management systems.

**Price: \$49.00**

**ISBN: 1-880124-04-1**

## **Z39.67-1993 Computer Software Description**

Computer software can sometimes defy description! Z39.67 gives guidelines for unequivocally describing software in advertising, on the packaging and carrier and labels, and on title screens.

**Price: \$ 45.00**

**ISBN: 1-880124-05-X**

## **Z39.58-1992 Common Command Language for Online Interactive Information Retrieval**

Useful to systems designers that want to specify a uniform command terminology, Z39.58-1992 describes nineteen non-proprietary command terms for use in online information retrieval systems. Defines the vocabulary, syntax, and operational meaning of the commands.

**Price: \$48.00**

**ISBN: 1-880124-03-3**

## **Scholarly Information and Standardization**

Eleven expert papers describe in detail national standardization activities in Canada, China, Japan, Korea, Singapore, the United Kingdom, and the U.S. Provides a wealth of information about global standards development, including coverage of international standardization work carried out by the ISO and TC 46, and ISO-IEC JTC 1. Edited by William E. Moen

**Price: \$40.00**

**ISBN: 1-880124-06-8**

## **Z39.47-1993 Extended Latin Alphabet Coded Character Set for Bibliographic Use (ANSEL)**

Character sets are a basic building block of automated information systems. Z39.47-1993 provides a table of coded values for the representation of characters of the extended Latin alphabet in machine-readable form for thirty-five languages written in the Latin alphabet and for fifty-one romanized languages.

**Price: \$40.00**

**ISBN: 1-880124-02-5**

**N**ISO Press –  
your source for  
standards for  
database management  
and design, information  
retrieval and networking

---

To order, contact NISO Press at: P.O. Box 338  
Oxon Hill, MD 20750  
(800) 282-NISO (6476)  
Fax: (301) 567-9553

Catalogs provided free on request.

**Reminder! Set-up a standing order and receive each new standard immediately on publication and get a discount.**

---

# The Rutgers Inventory of Machine-Readable Texts in the Humanities: Cataloging and Access

Annelies Hoogcarspel

*The Rutgers Inventory of Machine-Readable Texts in the Humanities was established in 1983 as a reference tool to help avoid duplication of effort for scholars and teachers in the humanities who want to use electronic texts in their work. The Inventory catalogers follow AACR2 and use the MARC format to provide bibliographic information about texts in all fields of the humanities, in any language, anywhere in the world, through the RLIN database. This article describes the information in the Inventory and some unresolved issues in relation to bibliographic control of electronic texts in the humanities.*

Since 1949, when Father Busa in Italy first started using a computer for work on Thomas Aquinas' writings, interest in computers as a tool for humanities research has been growing slowly but steadily. As more and more humanities scholars began working with computers, they soon found that the preparation of electronic versions of their primary source material was a tedious and painstaking task, and they began looking for existing electronic texts. Since there was not a single source of information about texts that had been converted to electronic format, or about the people who worked with electronic texts, information could only be obtained by word of mouth, through writing letters, or by asking around. Soon people found that they were duplicating effort, and it became clear that there was a need for a more reliable and accessible source of information about the existence and availability of electronic texts.

In the early eighties Marianne Gaunt at Rutgers University Libraries initiated the establishment of an online inventory of such texts, with financial support from the Council on Library Resources and the Mellon Foundation. In 1983, the first catalog record of the *Rutgers Inventory of Machine-Readable Texts in the Humanities* was created on RLIN, the online union catalog of the Research Libraries Group (RLG). RLIN was selected to hold the records because the materials were needed by scholars in research institutions. RLG was also involved in the adaptation of the MARC database format to hold this type of material. In 1991, work on the *Inventory* became one of the main activities of the newly established Center for Electronic Texts in the Humanities (CETH), with funding from the National Endowment for the Humanities.

This article describes the *Rutgers Inventory of Machine-Readable Texts in the Humanities* and the kind of information that is included in the *Inventory* records. It discusses how standard cataloging rules prescribe the information that must be included in a bibliographic description of a

text, ways in which these rules are enhanced for the *Inventory*, and the use of MARC fields for extensive notes and for access points. The next section covers RLIN and the new Eureka user interface and presents some examples of *Inventory* records. The final section addresses some of the unresolved issues involved in bibliographic control of and better access to electronic texts.

## □ Introduction to the *Inventory*

The *Inventory* is international in scope and covers all humanities areas: religion, literature, language, music, history, art, and philosophy. Primary source materials in these areas include literary works, historical documents, manuscripts, papyri, inscriptions, and transcriptions of speech and dictionaries and may be in any natural language. The words "electronic text" in this article refer to a transcription of this type of material into computer-readable form, character by character, including whatever additional information is needed to make the text useful.

An electronic text may be in one of two forms. One, often called a plain text or ASCII file, can be displayed, printed, or otherwise manipulated by whatever software an individual chooses to use or write. ASCII files can be easily copied, transmitted around the network, and updated and amended. The second form is indexed or organized for specific software and therefore usable only with that software (usually on CD-ROMs or disks). It is estimated that the large majority of existing electronic texts in the humanities are in ASCII format and are in the hands of individuals or research institutes who have compiled the texts for specific research projects. The remaining texts represent mostly commercially available products.

Electronic texts are created and archived under a variety of circumstances. A few are retained in electronic form when they are first written by the author, but the majority is either scanned in and converted from an image format to a text format or entered manually via a keyboard. People who subsequently edit the text and include markup are called editors and/or compilers in the context of the *Inventory*, depending on the amount of editing and/or encoding of the text. Other people and bodies receiving credit in *Inventory* records might be programmers who are involved with the preparation of text or software, publishers that might assume the responsibility of compilers, and

---

Annelies Hoogcarspel is Catalog Librarian, Center for Electronic Texts in the Humanities, Rutgers and Princeton universities. Her e-mail address is ceth@zodiac or ceth@zodiac.rutgers.edu.

archives that might also take on one or a combination of those responsibilities.

The sources of the texts in the *Inventory* are therefore varied: some are owned by individuals; others form part of the ARTFL database (American and French Research on the Treasury of the French Language); some are available, so far, from the Oxford Text Archive or the Norwegian Text Archive. A small but increasing number of commercially available scholarly items have also been included in the *Inventory*, such as the CD-ROM from CETEDOC, which contains works of the early Christian fathers, or the *New Oxford English Dictionary* on CD-ROM, or the CD-ROM with texts for use with the WordCruncher searching software.

Since 1983, more than 1,600 electronic texts and text corpora (collections) have been cataloged for the *Inventory*. Although a number of institutions are now cataloging the scholarly electronic materials they own for their online catalogs, the *Inventory* is the only effort to bring information about the existence and availability of these electronic texts together in one place, so that scholars and teachers in need of an electronic version of a text can establish, by consulting one database, if and where they might find it. However, because institutions such as Columbia University and the University of Virginia are now cataloging some of the same materials as CETH, the *Inventory* can benefit from the growing move toward cooperative cataloging as well.

There are now thousands of electronic texts all over the world. The *Inventory*, therefore, will grow. In the future, all commercially available electronic texts of interest to the scholarly community will be represented in the *Inventory*, as well as all texts available from the major text archives. In addition, CETH will continue to approach individuals for information about texts that they have and might be willing to make available to their colleagues. Work has also started on cleaning up the older records in the *Inventory*, some of which are, after all, more than ten years old.

The remainder of this article discusses in more detail what information is included in the *Inventory* records and how this information can be retrieved.

## Bibliographic Information in the *Inventory* Records

From the beginning of the *Inventory*, standard cataloging rules have been used for the bibliographic description of electronic texts. These rules are described in Chapter 9 ("Computer Files") of the *Anglo-American Cataloging Rules*, second edition, 1988 revision (AACR2).<sup>1</sup> They cover the description of all files that are encoded for manipulation by a computer and are therefore not especially geared toward the description of electronic texts as we define them. Some authors, such as Lynn Marko and Mary Ann

and Gerald B. Sheblé,<sup>2,3</sup> discuss problem areas and provide suggestions for improvement, but again concentrate primarily on software and data files rather than text files.

According to AACR2, information for the bibliographic description of any item is always taken from a chief source of information. For an electronic text, the chief source of information is the title screen. If there is no title screen, as is often the case, other internal evidence may be used (e.g., a main menu), and if this is not available, other sources include information from the physical carrier (e.g., a diskette) or its labels, from the publisher or creator (e.g., in documentation), or from the container of the carrier of the text file(s).

Cataloging for the *Inventory* is rarely done "with the item in hand," as the cataloging rules presuppose. The bibliographic information is therefore frequently derived from information supplied by the compiler of the electronic text on a survey or from documentation supplied by a text archive or database. Since it is mandatory in AACR2 to mention the source of the title information in a note, many records in the *Inventory* contain notes such as "Title from bibliography of texts supplied by contact person" or "Title from data abstract supplied by creator of file." Notes like these help identify the electronic text described.

Every title is followed by a "general material designator," which reads "[computer file]." This helps the individual who is looking for a certain text to distinguish between the electronic version and versions in other formats.

In addition to title information, statements of responsibility are taken from the same prescribed sources of information. They consist of the names of persons or bodies who are responsible for the intellectual content of the computer file. Most statements of responsibility in the *Inventory* contain a phrase indicating the type of responsibility of the individuals concerned, such as "compiled by."

The date of creation of the file is always recorded, even though it is sometimes only an estimate (e.g., "[198-?]" for a file probably created between 1980 and 1990). Information about the place of publication, the publisher, and the date of publication is recorded. Publication information in the *Inventory* is rare, however, because many texts are never published. To ensure that the legitimacy or authority of a particular text can be established from the record without publisher information, which often serves this purpose in the world of print materials, any other relevant information—for example, about an archive or distributor that makes the text available—is included in a note.

The file-characteristics area of a catalog record provides the prospective user of a text with some guidance on the type of file(s) and the amount of storage needed. For a text file, the cataloger states that the item consists of



"computer data" (as opposed to "computer program") and specifies how many files contain the complete work covered by the title and the number of bytes in each file. Both figures help the prospective user ascertain whether the electronic text is complete when received. The older records in the *Inventory*, however, do not always provide this information, especially if information was collected from questionnaires.

A description of the physical characteristics of the carrier of the electronic texts is required by AACR2. This description includes information about the number and type of physical units of the carrier, the dimensions of the carrier, and other details. This information is necessary for the prospective user to establish whether the text can be obtained in a form she or he can use. However, since most of the texts in the *Inventory* are available on many types of carriers or over a network, this information is less and less relevant. Exceptions to this are the commercially available items that are owned by CETH, for which detailed information about physical characteristics is included in the record. For texts that are available over the network, no physical characteristics are available, of course, but information on remote access is included in a note. Information about accompanying material (part of the physical description area) is very important, since the user might not have any other assistance in accessing a text or a text collection than through some written documentation.

AACR2 also provides a detailed list of information that might be specified in notes. Most often, notes clarify information provided in the bibliographic description of the text file or are used to provide additional points of access to a record. In the *Inventory*, notes provide information about the following:

- Source of the title
- Conditions under which the text is available
- Individual or institution from whom the text is available
- Print source of the electronic text
- Encoding features of the text
- Identification number of the text if it is part of one of the major text archives
- Available documentation
- Related publications

These notes, some of which are considered very important for assessing the usability of a text by a potential user, are discussed in more detail in the next section.

## Notes and MARC Fields

Cataloging information is transcribed into RLIN's adaptation of USMARC, RLINMARC. The MARC format

consists of several subformats, one of which is the MDF (Computer Files) format. Until the upcoming format integration (when all MARC fields can be used for any type of material), the MDF format specifies which MARC fields have been defined for cataloging computer files.

Most fields that are available in the MDF format can be used in at least one other format as well. Some fields, however, are especially well suited for information that is specific for electronic texts and very important to a user of the *Inventory* in assessing the usefulness of an item.

The 256 field is one of the small number of fields that is available only in the MDF format. It contains the file-characteristics information described in more detail above.

Many of the other fields that deserve special notice are note fields, in the 500 range. It is important to know that these fields are not indexed in RLIN and that many other library catalog search systems do not provide access to this information or do so only through limited keyword searching. It is therefore generally not possible to search for texts that contain certain text-encoding features or that will yield particular results when used with special software.

The 538 field (technical-details note) is used to describe the system requirements for using the electronic text. This information includes both central and peripheral equipment, as well as systems software. In this context, the 582 field (related-computer-files note) is used to provide information about related software that is needed to display or otherwise use the text. Another 538 field is used to describe additional file characteristics. These might include information about special features of the electronic version of the text, such as the occurrence of encoding, the level of encoding, and the type of features encoded. This information is of special importance to potential users of the texts in the *Inventory*, since it determines to a large extent what kind of research they can and cannot do with the text as it is and how much additional work they will have to do before they can use it for their purposes. The 538 field is also used when a text is available over the network to indicate that access is through a computer network.

A 537 field (source-of-data note) is used to transcribe information about the original print version of the text. Like other information about the provenance of a text (such as publisher or distributor information), this information helps assess the authoritativeness of the text.

In the 556 field (information-about-documentation note) full bibliographic information about any accompanying documentation is recorded. Although documentation is often not available, information about the availability of a manual or user guide may help a catalog user decide whether to make the effort to obtain the text. The 581 field (publications note) is sometimes used, especially with unpublished materials, to list publications that are based on the electronic text described in the record. The

530 field (additional-physical-form-available note) is used when CETH owns a copy of a commercial product that is also available in a different physical format than the one described in the 300 field for physical characteristics. The 506 field (restrictions-on-access note) contains information about the restrictions on access to the electronic text—for example, that a text is available for research, but not for teaching, purposes. Once again, information like this may help the user assess the usefulness of a particular text. The 510 field (references note), finally, is used to refer to the location of the text in a particular archive. It contains, for example, the unique text number assigned to a text in the Oxford Text Archive.

The general-note field, 500, has to be used to record any other information that is considered relevant in the context of the *Inventory*. Information about the special features of the software that is included with the text(s) is recorded in a 500 field, as well as names and addresses of contact persons for obtaining the text. The lack of sufficient note fields aggravates the problem that information in any of these fields cannot be searched in most online catalogs.

A new 856 field has been designed for information on remote access. This field allows for inclusion of electronic location and access information about electronic data resources. This field is the first step toward an environment in which catalog users can look up the bibliographic information about an electronic text they need and immediately move on to retrieve the text while they are looking at the bibliographic information. The information in this field, which is contained in a number of subfields and difficult for users to read, is in addition to a 538 field reading "Access through computer network."

## Access Points

Once the bibliographic description of the electronic text is complete, access points to the record are assigned. Rules for establishing name access are also provided by AACR2. Personal-name access points are assigned to records according to their forms in the Name Authority File (NAF) on RLIN, which ensures that an individual is always referred to by the same name and that no confusion between two individuals can occur. If the name is not in the NAF, it is established by the cataloger. In addition to access points that are customarily assigned in the *Inventory*, access is generally provided for an editor of the electronic version, or compilers and/or programmers involved with the electronic edition, as well as a text archive or database, a publishing company involved in the intellectual content of the work, and any others whose name might provide

additional access to the electronic text. Every record also has an access point for the *Inventory* itself.

In addition, the *Inventory* uses the *Library of Congress Subject Cataloging Manual* and the *Library of Congress Subject Headings* to provide subject access to the text files when appropriate.<sup>4,5</sup> *Genre Terms: A Thesaurus for Use in Rare Book and Special Collections Cataloging* is used to provide access to genre and form of the texts, since experience shows that many scholars who look for electronic texts might need "a corpus of English speech data," or "French novels."<sup>6</sup>

Access points are included in the regular fields for subject and name access. The 650 and 651 fields are used for topical and geographic subject headings from the *Library of Congress Subject Headings*, and the 690 field for terms from *Genre Terms: A Thesaurus for Use in Rare Book and Special Collections Cataloging*. Name access is provided through the 700 and 710 fields for personal and corporate names, and a local 797 field is used to provide access to the heading for the *Rutgers Inventory*. Additional title access is provided through 730 and 740 fields. Finally, another special MDF field, 753, provides an access point to the record for the computer system necessary to use the text. With the growing number of computer files, however, and the growing interchangeability between different computers of text files in particular, the use of this access point loses some of its importance.

## Retrieving Information from the *Inventory*

The description above deals mainly with the conventions relating to the input of information into the *Inventory*. A very important aspect, however, is the way in which users can retrieve the information they need from the *Inventory*.

The *Inventory* is available for searching at any RLG library or through a dial-up RLIN account.<sup>7</sup> More information on access to RLIN can be obtained from the RLIN Information Center at RLG. Inquiries about the *Inventory* can also be sent to CETH by those who do not have access to RLIN themselves.

In brief, the *Inventory* can be searched in the same way other RLIN searches are performed. This means that the "old" RLIN system interface can be used, as well as the newer Eureka search system, which has built-in help features and is therefore more user friendly. However, in Eureka all searches are performed across all formats, and it is therefore harder to locate and display computer file records only. Commands that do not work when the search result is too large ("all" on the headings screen) make it harder to do a fairly general search of the 1,600 records in the *Inventory* when searching the complete RLIN database

of more than 22 million items than it would be when searching the Computer Files (MDF file) of 33,000 records only in RLIN.

Under the old RLIN interface, the MDF file must be selected to search for *Inventory* records. It is highly recommended that in Eureka, which searches the complete bibliographic database, the type of materials be limited to MDF as well if the searcher is interested in *Inventory* records or electronic texts in general. An unlimited search would retrieve many unwanted records. An unfortunate feature of Eureka in our context is that even when the search is limited to a particular material, Eureka first displays a list of all headings in the RLIN database that contain the search terms specified, that is, not limited to MDF materials. However, once this headings screen is displayed (and the number of headings is not too big), typing the command "all" will display a screen that contains the titles of items in the MDF file only.

It is important to note that the MDF file contains all computer file records that are entered by catalogers at RLG libraries. These records include computer programs, data files, and electronic text files. Records of other libraries represent the holdings of a particular library, whereas the *Inventory* provides access to electronic texts that may be anywhere in the world. Under the old RLIN interface, *Inventory* records may be easily recognized by their Library of Congress library identifier, NJRP. In Eureka, each *Inventory* record displayed contains a note stating that "[t]his information is part of the *Rutgers Inventory of Machine-Readable Texts in the Humanities*"; a name access point for the *Inventory*; and, upon the user's typing the command "version njrp" (which may be abbreviated to "ver njrp"), the record's library identifier, NJRP. To list all *Inventory* records together, the name access point for the *Inventory* can be used. This name (*Rutgers Inventory of Machine-Readable Texts in the Humanities*) can be found in the corporate phrase index, which is searchable both under the old RLIN interface and in Eureka.

In general, records can be searched by author, title, and subject. For example, a search for a text by the author Ernest Hemingway in Eureka, limited to computer files, would retrieve the following record:

TITLE: [Thirty-nine short stories and essays] [computer file]/ compiled by students of the Catholic university.  
PUBLISHED: 1985.  
PHYSICAL DETAILS: Computer data (5 files)  
OTHER AUTHORS: Restelli, Raffaella.  
Universita Cattolica del Sacro Cuore.  
Gruppo Interdisciplinare di Ricerce per la Computerizzazione dei Segni dell'Espressione.  
Rutgers Inventory of Machine-Readable Texts in the Humanities.  
OTHER ENTRIES: Hemingway, Ernest,

1899-1961. Short stories. Selections. 1985.  
Fitzgerald, F. Scott (Francis Scott), 1896-1940. May day. 1985. Fitzgerald, F. Scott (Francis Scott), 1896-1940. Alcoholic case. 1985.  
Fitzgerald, F. Scott (Francis Scott), 1896-1940. Cut glass bowl. 1985.  
Fitzgerald, F. Scott (Francis Scott), 1896-1940. Diamond as big as the Ritz. 1985.  
Eliot, T. S. (Thomas Stearns), 1888-1965. Cocktail party. 1985.  
Eliot, T. S. (Thomas Stearns), 1888-1965. Confidential clerk. 1985.  
Orwell, George, 1903-1950. Essays. Selections. 1985.  
Lawrence, D. H. (David Herbert), 1885-1930. Prussian officer. 1985.  
SUBJECTS: English language-Written English-Databases.  
Short stories-United States-20th century.  
Essays-United States-20th century.  
NOTES: In American English.  
Title supplied by cataloger based on information provided by contact person.  
File size: 78,809, 42,296, 48,374, 47,216, 9,044 words. Data collected to build concordances and an initial English machine dictionary to lemmatize other texts, by students of the Gruppo Interdisciplinare di Ricerce per la Computerizzazione dei Segni dell'Espressione of the Universita Cattolica del Sacro Cuore.  
CONTENTS: file 1. Short stories between 1923/1925 and 1933/1938 / E. Hemingway - file 2. May day. An alcoholic case. The cut glass bell. The diamond as big as the Ritz / F.S. Fitzgerald - file 3. The cocktail party. The confidential clerk / T.S. Eliot - file 4. Six essays / G. Orwell - file 5. The Prussian officer / D.H. Lawrence.  
NOTES: Raw texts and concordances available for researchers free of charge. Data provided on magnetic tape; includes description of encoding system.

It is possible to locate this particular selection of short stories by Hemingway because access has been provided on an analytic level, listing all authors in the text collection with story titles where known. The *Inventory* would be most useful, of course, if all texts received this kind of analytic treatment. A number of CD-ROMs, such as the ICAME CD-ROM, are cataloged with separate descriptions for each of the corpora contained on them.

A title keyword search for a copy of *The Paston Letters* using Eureka yields a list of titles small enough so that someone looking for any format may discover the computer file among all other titles listed; it is number 22 in a list of eighty. However, users would most likely be looking for an electronic text in particular at the present stage of electronic text usage among humanities scholars, and in that case this text would stand out on its own.

TITLE: Paston letters and papers of the fifteenth century [computer file] / compiled by Norman Davis.  
 PUBLISHED: [ca. 1984]  
 PHYSICAL DETAILS: Computer data (3 files : ca. 532, 550, 172 Kb).  
 OTHER AUTHORS: Davis, Norman, 1913- Rutgers Inventory of Machine-Readable Texts in the Humanities. Oxford Text Archive.  
 SUBJECTS: Paston family. England-Social life and customs-Medieval period, 1066-1485. Letters-England-15th century.  
 NOTES: In Middle English. Title from t.p. of source text. Available for teaching and research purposes. Transcribed from: Paston letters and papers of the fifteenth century / edited by Norman Davis. Part 1. Oxford : Clarendon Press, 1971. Contains markup characters. This information supplied to the Rutgers Inventory of Machine-Readable Texts in the Humanities.  
 REFERENCES: Oxford Text Archive 0395.

The record retrieved shows that this particular text is available from the Oxford Text Archive, and it lists the unique text number.

A subject search to retrieve records for corpora in the *Inventory* would have to include the genre term "anthologies," which is used for lack of a better term in *Genre Terms: A Thesaurus for Use in Rare Book and Special Collections Cataloguing*. Searching RLIN under the old RLIN interface, the first screen contains the following titles:

- 1) THE LONDON-LUND CORPUS [COMPUTER FILE] / In: ICAME collection of English language corpora / [compiled by the] Norwegian Computing Centre for the Humanities. c-9114 NJRP92-D26
- 2) Lancaster-Oslo/Bergen corpus THE LOB CORPUS [COMPUTER FILE] / In: ICAME collection of English language corpora / [compiled by the] Norwegian Computing

Centre for the Humanities. c-9114 NJRP92-D25

- 3) THE KOLHAPUR CORPUS [COMPUTER FILE] / In: ICAME collection of English language corpora / [compiled by the] Norwegian Computing Centre for the Humanities. c-9114 NJRP92-D24
- 4) THE HELSINKI CORPUS OF ENGLISH TEXTS [COMPUTER FILE] / In: ICAME collection of English language corpora / [compiled by the] Norwegian Computing Centre for the Humanities. c-9114 NJRP92-D23
- 5) Standard corpus of present-day American English A STANDARD CORPUS OF PRESENT-DAY EDITED AMERICAN ENGLISH [COMPUTER FILE] / c-9114 NJRP92-D22

The complete record for the London-Lund corpus:

The London-Lund corpus [computer file] / collected and transcribed under the direction of Randolph Quirk ; computerized under the direction of Jan Svartvik. Computer data (ca. 59 files : 65 MB). Several directories on 1 computer laser optical disk : 4 3/4 in. System requirements: IBM PC or compatible or Macintosh or UNIX system; hard disk; CD-ROM drive. ASCII text files. Orthographic transcription with detailed prosodic marking: stress, intonation, pauses, etc. "Original text versions for MS-DOS, Macintosh and Unix. An edited version indexed by WordCruncher 4.4 and TACT for MS-DOS and Free Text Browser for Macintosh."-info.w51 file. Title from info.w51 file. Collected and transcribed at the Survey of English Usage, University College of London, and computerized at the University of Lund.-info.w51 file. Summary: "The SEU corpus contains 200 samples or "texts," each consisting of 5,000 words, for a total of one million words. The texts were collected over the last 30 years, half taken from spoken English and half from written English."-The London-Lund corpus of spoken English / Sidney Greenbaum & Jan Svartvik. In The London-Lund corpus of spoken English : description and research / ed. by Jan Svartvik. Lund : Lund University Press, 1990. (Lund Studies in English ; 82). Available for research (P.T.O.) purposes only. Available for demonstration purposes at the Center for Electronic Texts in the Humanities.

This information is part of the Rutgers Inventory of Machine-Readable Texts in the Humanities.

In: ICAME collection of English language corpora / [compiled by the] Norwegian Computing Centre for the Humanities. - Bergen, Norway : The Centre, 1991.

1. English language-Spoken English-Databases. 2. English language-Written English-Databases. 3. Anthologies-Sweden-20th century. 4. Anthologies-Great Britain-20th century. I. Svartvik, Jan. II. Quirk, Randolph. III. Lunds Universitet. IV. University College, London. V. Rutgers Inventory of Machine-Readable Texts in the Humanities. VI. Title.

In general, it is advisable to use the keyword-searching capabilities of either the old RLIN indexes or the new Eureka index, unless the subject heading lists or name authority files are consulted first and a complete name or subject search is prepared based on the headings in those lists.

### Some Remaining Issues in the Bibliographic Control of Electronic Texts

Several unresolved issues emerge from a more detailed look at the procedures involved in creating and searching the *Rutgers Inventory of Machine-Readable Texts in the Humanities*.

The inconsistency and incompleteness of the sources of bibliographic information continue to be a problem. Many texts, both those commercially available and those obtained from individuals or institutions, do not have title screens. Those that do often have incomplete or ambiguous information on them. Second, more often than not, a reference to the original print version of a text is lacking in a cataloging record, because this information is so difficult to obtain. The information is vital, however, to the scholar. Third, cataloging without the item in hand is a major problem for many libraries cataloging electronic texts that they do not own, because of a lack of good documentation from the source of the text. This also holds true in those cases where a cataloger has the item, but does not have the hardware and/or software to actually look at its contents. Documentation is also important for users to be able to use the capabilities of a text and/or its accompanying software to the fullest extent.

The Text Encoding Initiative (TEI) guidelines are a step toward the resolution of these problems.<sup>8</sup> The TEI is a major international project established in 1988 to develop guidelines for the preparation of machine-readable texts for scholarly research and to satisfy a broad range of uses by the language industries more generally. Sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing, the guidelines were developed with input from a wide range of potential users in different discipline areas and use the Standard Generalized Markup Language (SGML) for encoding. These guidelines also include a chapter on the description of a header that precedes the electronic text and that can serve as the primary source of information for cataloging electronic texts. In addition, this header contains information on the electronic text, its source, its encoding, and its revisions. In short, it incorporates all the information catalogers need to follow the rules in chapter 9 of AACR2, which would make it an excellent chief source of information for catalogers. In addition, the header would contain other information that catalog users need when they are looking for electronic texts, but that is now often lacking in catalog records. For example, it would resolve the issue of the existence of different versions of the same text. It would also provide information about the markup contained in a text.

Access to electronic texts for humanities research would be improved if more appropriate ways of indexing them were available. Form and genre access were never part of the Library of Congress's subject heading treatment, and fiction in general does not receive any subject access. The genre term thesaurus is not entirely satisfactory because it covers rare book and archival materials, and, as we have seen in this article, modern humanities computing terms such as "corpus" are not represented. In addition, some of the information that we have seen is very important to users of electronic texts is buried in nonsearchable note fields.

A final problem is related to the copyright questions that surround all use of electronic resources. Some owners of electronic texts do not want to share information about them, because they have reason to believe that they might run into copyright problems if it becomes known that they have those texts. Although this might sometimes be the case, information about the availability of texts in cataloging records may be included to indicate that texts are not available for general public use. At the same time, including cataloging for the text would allow scholars with a serious interest to establish whether a copy of the text they need exists.

These issues related to the bibliographic control of and access to electronic texts can only be resolved when



# Poststructuralism and the ARTFL Database: Some Theoretical Considerations

Mark Wolff

*In order to attract new scholars to computers in the humanities, some researchers suggest that poststructuralist theories should inform how one queries textual databases. This paper argues that if we adopt a poststructuralist methodology to analyze texts with computers, we must recognize the assumptions that motivate poststructuralist theories. Using such theories to find an ultimate meaning in literature would entail contradictions because these theories deny the possibility of an ultimate meaning in literature. Although computers can be used effectively in conjunction with current critical theory, they will never uncover the one true meaning of a text.*

The fact that present online text databases offer seemingly limitless opportunities to analyze literature in comprehensive ways has prompted much optimism about the future of computers in the humanities. Indeed, large corpora of texts immediately available in their entirety with a few strokes on the keyboard represent a remarkable achievement in applying computer technology to disciplines in the liberal arts. By rendering texts written in natural-language machine-readable, students of literature can let the computer do tasks that even with an authoritative concordance would be tiresome and time-consuming. Thematic studies that required long hours of reading and rereading in order to observe how certain words express a particular idea can now be executed in less than a minute. This radical reduction in the time necessary to work out a literary analysis enables one to pursue more ambitious studies that can treat many themes at once. In addition to thematic studies, the development of sophisticated morphological analyzers is opening the way to systematic linguistic and semantic investigations into the imperceptible shifts and transformations of language that occur within a text or over time through many texts. Not only can we see how ideas are developed through language, but we can also observe how language itself operates to produce meaning. Given the exponential rate of advances in computer technology over just the past ten years and the promise of even more amazing things to come, it is no wonder that the possibilities for computerized literary studies appear endless.

Unfortunately, most of the academic community does not share this enthusiasm. In a recent essay, Mark Olsen has addressed the lukewarm response of the academic community to computers in their midst, identifying two reasons that scholars seem reluctant to embrace the new technology.<sup>1</sup> First, previous literary studies employing computer methodologies have appealed only to researchers who have already adopted these methodologies

in their own work. Instead of attracting new people to accept and use the technology, these individuals have ended up forming an isolated discipline complete with journals and conferences that nobody else reads or attends. The reason for this is that the wider academic community does not appreciate the conclusions drawn from the results of computerized literary analyses: as far as the majority is concerned, these analyses do not really tell us anything we did not know without computers. Second, those who use computers to read literature have fallen out of touch with recent developments in literary criticism, namely poststructuralism. While the rest of academia was wrestling with semiotics, psychoanalysis, reader-response theory, and deconstruction, computer buffs continued to crunch words to uncover meanings without questioning what "meaning" meant. As a result, those who have been devoted to computers during the past ten to fifteen years find themselves removed from mainstream literary debates because their work has not contributed to current critical theory and has more or less ignored it.

In order to demarginalize computer literary studies, Olsen proposes that those who use computers to study literature incorporate poststructuralist theories into their research models:

The most surprising element of this problem is the relatively limited acceptance of critical models that stress intertextuality and sign theory among those interested in computer analysis of text. This is not to say that the structuralist and post-structuralist critical theories are not talked about in relationship to computer-assisted textual research, but that few systematic attempts have been made in this area, combining software, data selection, and theory in interesting ways.<sup>2</sup>

Developing an interface between poststructural theories and computer technologies would benefit both approaches to the study of literature. On the one hand, poststructuralists would be able to test their theories empirically by exploring such phenomena as intertextuality and sign systems. Because of their high speed and access to large amounts of text, computers could become for them an invaluable tool. On the other hand, those who wish to pursue computer analyses of literature would find new methodologies for their research, which has up to now been following the same model with little success. Poststructuralism would offer computer text analysis the theoretical backbone it has been lacking. This proposition fits in with the current interest in interdisciplinary studies, and

---

**Mark Wolff** is the coordinator of data entry for the ARTFL Project and a graduate student of French literature at the University of Chicago.

for that reason alone computers should find wider acceptance among scholars in general. As researchers continue to look for new perspectives on cultural phenomena, they will turn to areas of study that have yet to be exploited. The combination of computers and literary studies is simply inevitable.

The ARTFL database is one example of how computers have already made an impact on literary studies. ARTFL, or the Project for American and French Research on the Treasury of the French Language, is a collaborative effort undertaken by the University of Chicago and the Institut National de la Langue Française (INaLF) of the Centre National de la Recherche Scientifique (CNRS). In existence for more than ten years, it represents one of the longest lasting and most successful international joint projects in the humanities and is becoming one of the basic tools of scholarly research in literature and history. Its database consists of nearly two thousand texts from the sixteenth to the twentieth centuries. Genres such as novels, poetry, correspondence, journals, histories, and treatises are represented, providing scholars a wide variety of texts from which to choose. ARTFL's current text-retrieval software, Philologic 2.0, allows scholars to conduct sophisticated word searches and even statistics in a user-friendly environment. These searches include tracing concepts in various authors over time, comparing how authors differ in their use of words and concepts, or simply discovering which authors employ certain types of language and in what context. ARTFL is constantly enlarging its database and improving its software, indicating the remarkable potential for computer-assisted text analysis.<sup>3</sup>

Although there is nothing stopping us from hacking intricate UNIX code after reading Roland Barthe's *S/Z* (this may in fact be an appropriate response), we should be careful about how we do both. One of the difficulties with interdisciplinary studies is that when one tries to do two different things at once, one often succeeds at neither. Computer programmers would think it naive if someone told them a computer can be made to do anything, because programmers must invent and tailor their tasks around the capabilities of the computer apparatus. The same must be said about poststructuralist theories. To claim that critical theory can solve the endemic problems of literary analysis with computers risks putting too much faith in critical theory. Poststructuralist theories have apparatuses such as terminologies and methodological procedures; these elements cannot adapt themselves effortlessly to analyzing large chunks of text with a computer. In order to use computers effectively in conjunction with current literary theory, both disciplines must be practiced well.

Perhaps the most serious error that computer specialists can make in applying poststructuralist theories to literary research is assuming that these theories are better means for attaining the same goal that has always charac-

terized computer research in the humanities: the irrefutable meaning of a text. This is not to say that all computer researchers believe the same thing about the meaning of and in literature. Some investigate single texts in order to identify the deep structures that build and frame a text's particular meaning. Others look at large corpora of texts to trace the structures of intertextuality, i.e., the ways in which a text constantly draws from and refers to other texts in a web of historical and contextual relations that readers normally cannot perceive because the web is so vast. Despite the conceptual differences between these two perspectives on literature, both perspectives usually assume that there is a unique signified (in Saussurean semiology) or representamen (in Peircean semiotics) behind every element of text. The ultimate source for "meaning" differs between micro- and macro-perspectives on literature, the former assigning it to the Author or the intended Reader (the *destinataire*), the latter to cultural systems in a specific time and place, but both aspects lead their proponents to find the one interpretation that explains and justifies how a text *must* be read. Jean-Jacques Thomas makes explicit this assumption about the ultimate goal of literary research. Citing Riffaterre's *Essais de stylistique structurale*, Thomas argues that

in the end, having read and analyzed a given text, all readers will "necessarily be in agreement on the text's single meaning" . . . this declaration of principle is, in [his] view, [Riffaterre's] theoretical mainstay, founding, as it does, a kind of hermeneutics of discovery . . . animated by the belief that all careful readers will arrive at one and the same true and universal interpretation of the text.<sup>4</sup>

Now if one agrees with this principle, there is no reason that literary research should not proceed according to the assumptions outlined above. However, Riffaterre's book was published in 1971 and must be considered a structuralist approach to literature or even an example of the New Criticism of the 1950s. To hold these assumptions and look to poststructuralist theories for a computer research model is to overlook the assumptions that motivate later critical efforts. Although poststructuralism is a catch-all term that encompasses many different theoretical orientations, they all share a common mistrust of the "true and universal interpretation of the text." Theories such as Riffaterre's attempt to legislate a method because the act of reading the text must be controlled by standard procedures in order to obtain unified results. This approach assumes that ultimately all readers will subordinate their personal interpretations for the sake of a universal one. Obviously, as literary scholars we are engaged in a praxis that produces knowledge, but must our interpretations of texts necessarily contribute to a unified body of knowledge? If individual scholars insist on their own readings of texts, does that mean they are no longer pursuing knowl-



edge? Antony Easthope identifies two competing paradigms for literary studies that derive from the researcher's assumption about knowledge:

If the work seems unified, that confirms its identity independent of the reader so it can be treated as "there," an object within the empiricist scenario; but if the text cannot be seen as unified, if it consists of a plurality of meaning which is potentially infinite, then it can only be the reader working according to a theory of reading which can provide a unification by nature never more than provisional. The schema [of the paradigm for reading] shifts from:

Reader ———> Text (= Author)

to one with a double-headed arrow in which text and reader interact dialectically:

Reader ———> Text<sup>5</sup>

In the case of intertextual studies with large corpora of texts, the quest for structures of language that determine meaning would fall under the first paradigm, which we could modify as:

Reader ———> Text (= Intertextuality)

The upshot of this distinction is to demonstrate that structuralist and poststructuralist approaches to literature entail opposing agendas that are not easily reconciled. The decision to use poststructuralist theories must imply a prior decision to at least suspend the belief that texts have unique meanings that can be extracted with the right tools. If one believes that texts have unique meanings *and* wants to use poststructuralist theories, contradictions will ensue.

In his essay "The Death of the Author," Roland Barthes provides a critique of the logocentric empiricism that other poststructuralist theories reject as a basis for their interpretive practices. For him, a text does not contain "a single 'theological' meaning (the 'message' of the Author-God) but a multidimensional space in which a variety of writings, none of them original, blend and clash."<sup>6</sup> This assertion would negate any attempt to analyze a single text in the hope of uncovering a hidden meaning or structure that eludes the attention of traditional readers. One might think, however, that the "multidimensional space" where texts reside can be systematically analyzed in order to map out empirically all the structures that produce meaning. The origin of the text's meaning would not be the "Author-God" but this "multidimensional space," which is intertextuality. Barthes denies this possibility:

Once the Author is removed, the claim to decipher a text becomes quite futile. To give a text an Author is to impose a limit on that text, to furnish it with a final signified, to close the writing. Such a conception suits criticism very

well, the latter then allotting itself the important task of discovering the Author (or its hypostases: society, history, psyché, liberty) beneath the work: when the Author has been found, the text is "explained"—victory to the critic. . . . In the multiplicity of writing, everything is to be *disentangled*, nothing *deciphered*; the structure can be followed, "run" (like the thread of a stocking) at every point and at every level, but there is nothing beneath: the space of writing is to be ranged over, not pierced; writing ceaselessly posits meaning ceaselessly to evaporate it, carrying out a systematic exemption of meaning. In precisely this way literature (it would be better from now on to say *writing*), by refusing to assign a "secret," an ultimate meaning, to the text (and to the world as text), liberates what may be called an anti-theological activity, an activity that is truly revolutionary since to refuse to fix meaning is, in the end, to refuse God and his hypostases—reason, science, law.<sup>7</sup>

The practice of interpreting texts is, in the poststructuralist paradigm, not so much a subjective exercise in reading that refuses ultimate meaning but a rewriting of texts within particular critical discourses that meet specific political agendas. Marxist and feminist critiques of texts exemplify this practice, but in fact any interpretive practice carries with it certain a priori beliefs that inform assertions to be proven, not hypotheses to be tested. Those who insist that a text has a unique meaning perform the same kind of interpretive practice as poststructuralists; the difference between them is that poststructuralists acknowledge the contingencies of their interpretations.

As long as researchers resist pretensions to scientific truth, computers can help unravel possible meanings that would be too difficult otherwise. One way to use computers effectively without claiming universality is to explore discursive formations that underlie *how* we read texts. The emphasis of such an approach would be to locate words, phrases, and syntactic constructions that produce meanings in a particular instance of reading. In order to demonstrate the strengths and weaknesses of intertextual computer analysis conducted in this way, I will use Naomi Schor's reading of George Sand and realist discourses to query the ARTFL database about the relationship of idealism to realism in nineteenth-century French literature.

In her essay "Idealism in the Novel: Recanonizing Sand," Schor's point of departure is the decline of George Sand's status as an author in the literary canon of late nineteenth-century France. Schor's thesis is that

Sand's spectacular aesthetic devaluation cannot be ascribed in any simple terms to her gender; it is not because Sand was a woman, rather because (like so many other women authors) she is associated with a discredited and discarded representational mode that she is no longer ranked among the canonic authors.<sup>8</sup>

This representational mode is idealism, whereby the world and its concrete realities are portrayed positively through

**Table 1**  
Frequency of Words Matching "idE[aeo].\*\*"

WORD	PER 1ST CORPUS		PER 2ND CORPUS		WORD	PER 1ST CORPUS		PER 2ND CORPUS	
	(date = 1810-1849)	PER 1,000	(date = 1850-1889)	PER 1,000		(date = 1810-1849)	PER 1,000	(date = 1850-1889)	PER 1,000
Total	12,658	0.791434	11,863	0.697940	idéalité	16	0.001000	14	0.000824
idéal	480	0.030012	1,141	0.067129	idéalités	6	0.000375	2	0.000118
idéale	100	0.006252	173	0.010178	idéals	4	0.000250	4	0.000235
idéalement	8	0.000500	15	0.000883	idéaux	0	0.000000	6	0.000353
idéales	28	0.001751	52	0.003059	idee	0	0.000000	1	0.000059
idéalisai	1	0.000063	0	0.000000	idée	6,590	0.412036	6,439	0.378828
idéalisait	0	0.000000	1	0.000059	ideen	1	0.000063	0	0.000000
idéalisant	1	0.000063	1	0.000059	idéenne	1	0.000063	2	0.000118
idéalisation	7	0.000438	13	0.000765	idéens	1	0.000063	0	0.000000
idéalisations	1	0.000063	0	0.000000	idéer	1	0.000063	0	0.000000
idéalisatrice	0	0.000000	1	0.000059	idées	5,189	0.324439	3,829	0.225273
idéalisée	9	0.000563	8	0.000471	idéesd'	0	0.000000	1	0.000059
idéalisé	5	0.000313	7	0.000412	ideo	1	0.000063	2	0.000118
idéalisée	6	0.000375	12	0.000706	idéo	0	0.000000	1	0.000059
idéalisées	0	0.000000	1	0.000059	idéographique	0	0.000000	3	0.000177
idéalisent	0	0.000000	1	0.000059	idéologie	18	0.001125	5	0.000294
idéaliser	6	0.000375	16	0.000941	idéologies	1	0.000063	0	0.000000
idéalisés	2	0.000125	4	0.000235	idéologique	3	0.000188	1	0.000059
idéalisiez	0	0.000000	2	0.000118	idéologiques	1	0.000063	0	0.000000
idéalisme	93	0.005815	36	0.002118	idéologistes	1	0.000063	0	0.000000
idéalisons	0	0.000000	2	0.000118	idéologue	7	0.000438	10	0.000588
idéaliste	48	0.003001	33	0.001942	idéologues	9	0.000563	8	0.000471
idéalistes	13	0.000813	16	0.000941					

melioristic devices such as hyperbole in order to elevate the image of the world above the problems of contemporary life. As Schor points out, realism also sought to improve contemporary life, albeit through a negative practice of representing the baser aspects of reality; realism thus falls under the rubric of idealism. The devaluation of writers such as Sand occurs when the difference between the two modes of idealism is linked to sexual difference: negative realism is associated with the masculine and valorized over positive idealism, which is associated with the feminine.<sup>9</sup> Schor's claim that realism became a privileged discourse over idealism would seem to invite a systematic analysis of the ARTFL database in order to support or refute her claim. This, however, would not be a simple task. Realism and idealism do not distinguish themselves by *what* they say but *how* they say it. Table 1 compares the frequency of words matching the pattern "idE[aeo].\*\*" during the periods of 1810-49 and 1850-89.<sup>10</sup> Although words matching this pattern are more generally frequent in the

earlier period, keywords such as *idéalisme*, *idéaliste*, *idéalité*, and even *idée* are more prevalent in the later period. In table 2 there is no significant difference in the total frequency of words matching the pattern "rE[ae].\*\*" during the periods of 1810-49 and 1850-89, but certain words such as *réalisme*, *réaliste*, and *réalité* occur more frequently during 1850-89 while *réel* and its variants (*réels*, *réelle*, *réelles*, *réellement*) are used more often during 1810-49. Thus a statistical description of the frequency of words related to the real and the ideal during the nineteenth century is inconclusive. Exploring the gender aspect of this question is even more problematic, given the ubiquity of words such as *homme* and *femme*. In order to see whether realism and idealism are constructed differently in the discourses of the nineteenth century, we have to read how these words are being used. Is idealism talked about cynically in the latter half of the nineteenth century? What do people mean by the real before 1850? Computers cannot answer these questions because they require us to interpret their results. The piv-

**Table 2**  
Frequency of Words Matching "rE[ae]l.\*"

WORD	PER 1ST	PER 1,000	PER 2ND	PER 1,000	WORD	PER 1ST	PER 1,000	PER 2ND	PER 1,000
	CORPUS	(date = 1810-1849)	CORPUS	(date = 1850-1889)		CORPUS	(date = 1810-1849)	CORPUS	(date = 1850-1889)
Total	3,770	0.235717	3,822	0.224861	réaliserait	7	0.000438	9	0.000530
real	1	0.000063	0	0.000000	réaliseras	1	0.000063	0	0.000000
réal	1	0.000063	0	0.000000	réalisèrent	4	0.000250	6	0.000353
reale	1	0.000063	10	0.000588	réaliserez	2	0.000125	1	0.000059
réales	1	0.000063	0	0.000000	réaliserait	1	0.000063	0	0.000000
réalisa	10	0.000625	16	0.000941	réaliseront	2	0.000125	0	0.000000
réalisable	6	0.000375	14	0.000824	réalises	1	0.000063	0	0.000000
réalisables	2	0.000125	2	0.000118	réalisés	15	0.000938	17	0.001000
réalisai	2	0.000125	1	0.000059	réalisez	4	0.000250	0	0.000000
réalisaient	9	0.000563	8	0.000471	réalisme	8	0.000500	66	0.003883
réalisais	2	0.000125	0	0.000000	réalismes	0	0.000000	1	0.000059
réalisait	25	0.001563	37	0.002177	réalisent	1	0.000063	0	0.000000
réalisâmes	1	0.000063	0	0.000000	réalisaient	1	0.000063	0	0.000000
réalisant	15	0.000938	10	0.000588	réalisait	1	0.000063	0	0.000000
réalisas	0	0.000000	1	0.000059	réalisait	1	0.000063	0	0.000000
réalisassent	1	0.000063	1	0.000059	réalisons	0	0.000000	2	0.000118
réalisât	4	0.000250	5	0.000294	réaliste	3	0.000188	31	0.001824
réalisateur	0	0.000000	1	0.000059	réalistes	8	0.000500	25	0.001471
réalisation	69	0.004314	74	0.004354	réalité	824	0.051520	1,126	0.066246
réalisations	1	0.000063	6	0.000353	réalités	162	0.010129	137	0.008060
réalise	40	0.002501	50	0.002942	reality	1	0.000063	0	0.000000
réalisé	68	0.004252	80	0.004707	réel	486	0.030387	443	0.026063
réalisée	34	0.002126	33	0.001942	réele	1	0.000063	0	0.000000
réalisées	15	0.000938	30	0.001765	réelection	1	0.000063	0	0.000000
réalisent	11	0.000688	21	0.001236	réelement	1	0.000063	0	0.000000
réaliser	221	0.013818	218	0.012826	reelle	1	0.000063	0	0.000000
réalisera	12	0.000750	13	0.000765	réelle	479	0.029949	401	0.023592
réaliserai	2	0.000125	2	0.000118	réellemen	1	0.000063	0	0.000000
réaliseraient	1	0.000063	6	0.000353	réellement	874	0.054646	708	0.041654
réaliserais	2	0.000125	2	0.000118	réelles	152	0.009504	108	0.006354
					réels	171	0.010692	100	0.005883

otal moment in any textual analysis is the choice readers make when confronted with an element that needs to be explained. Even though we can read more text more systematically with computers, we must still contend with our "horizon of expectations," a preexisting frame of reference that governs how we interpret texts according to subjective perceptions.<sup>11</sup>

Perhaps one way to use computers effectively in textual analysis is to see how the text is able to manipulate

how we read. Intertextuality from this angle would not be a static system of fixed signifiers but rather openings in the text that compel the reader to participate in the production of meaning. Jonathan Culler has suggested that a linguistic model might be a fruitful way to explore intertextuality. Natural language as it is spoken and written often includes presuppositions that require the listener and reader to place themselves in imaginary frameworks with only a few details to guide them. Usually context alone will determine

date = 1810-1889: 447 document(s); Philologic: 2.0

Table 2

Library of Words Introducing Presuppositions

		Keyword in context:
1/78, done		
St1	Allem 229	resser , et cependant quand on veut qu' elle soit joué
St1	Allem 229	u' elle soit jouée , quand on lui dit lève - toi et
Lmrtn	MedPo 34	se reconnoît plus au moment du réveil . hélas ! tel
LsCss	MeStH 746	insinuer sans cesse . quant à moi , je confessais de b
Jbrt	PenCo 175	e la lumière , qui , lorsqu' il leur vient une idée a
Jbrt	PenCo 175	prescrivait #Buffon , quand il définissait le génie l
Jbrt	PenCo 175	ouvait des diamants , quand on en brisait l' enveloppe
MndBrn	Jour2 52	valeur réelle ; mais quand l' homme est éclairé et r
MndBrn	Jour2 95	le sait et le voit . lorsque je sens ma nullité , un
MndBrn	Jour2 189	n elle à son insu au moment de la création , il est c
MndBrn	Jour2 189	ou les connaît , et quand ou comment a lieu cette ape
MndBrn	Jour2 231	é dans ma journée , quand j' ai travaillé avec un su
MndBrn	Jour2 271	ait désintéressé ; quand nous cherchons la vérité
MndBrn	Jour2 285	sentiment ; ou même lorsque je nourris ce sentiment a
MndBrn	Jour2 360	ce qui est en nous . quant à l' idéal que nous a
Csn	CHiPh 115	; mais que sera - ce lorsque l' on s' interdit même c
Csn	CHiPh 115	tour à la vérité , lorsqu' on omet complètement la
Csn	CHiPh 117	ls de nos idées ; et quand viendra tardivement l' obse
Csn	CHiPh 292	elle - même ; car , quand même vous avez l' idée d'
Csn	CHiPh 307	t incontestable que , lorsque vous parlez du livre en g
Csn	CHiPh 307	ire , je demande si , lorsque vous parlez de l' espace
Csn	CHiPh 322	cet égard . ainsi , lorsqu' à la fin du second livre
Csn	CHiPh 419	bstraction moi . mais quand même cela serait possible
BlzCh	PeaCh 224	en porter ici - bas . quand il fut assis dans son faute
Ndr	FeeMi 109	é . à compter de ce moment , je pris le ferme dessein
BlzCh	Corr1 563	hions les cheveux par moments , faute de pouvoir réali
Krr	SoTil 116	'hui , #Magdeleine , quand ton amour , si réellement
Lmrtn	SoVoO 331	losophe , est dans un moment de dérouté ; elle n' a ni
Vgn	SerGM 165	turel et vrai dans ce moment - là , il ne songeait poi
Gzln	NotCh 23	es deux cousines , le moment ou\ les jeunes filles , fa
BlzCh	CesBi 33	P 33 #Paris . aussi , quand #Ursule l' abandonna , fut
Sl	MemDi 332e	e les appeler . à ce moment une idée étrange vint à
Lrx	Human 60	es du christianisme , quand les trois termes divins de
Lrx	Human 60	e de sortir , surtout lorsque , joignant le stoïcisme
Lrx	Human 264	sont identiques . et lorsque #Dieu a ainsi fait l' hom
Lrx	Human 603	par le nombre 3000 , quand , le déluge ayant détruit
Prdhn	Propr 138	ments les plus faux , quand ils P 139 ont pour base des
SnnCr	Oberm 104	blé de douleurs . et lorsque l' imagination reportée
SnnCr	Oberm 158	éveiller amèrement quand tout repose , et chercher d
SnnCr	Oberm 158	ercher des ténèbres quand la terre fleurit ; pour n'

Figure 1  
KWIC Listing of Words Introducing Presuppositions

whether a sentence contains a presupposition, but certain words often serve to introduce presuppositions.

For example, consider the following sentence taken from the second volume of Maine de Biran's *Journal*:

Quant à l'*idéal* que nous aimons, comme il est dans ce système l'*ouvrage de notre esprit*, en aimant cet idéal, nous n'aimons point en effet ce qui est hors de nous; mais aussi nous n'aimons, nous n'embrassons rien de réel.<sup>12</sup>

This sentence was located in the ARTFL database with a co-occurrence search for the patterns "idE[aeo].\*", "rE[ae]l.\*" and "quant." I specifically wanted to find sentences containing a "quant" phrase that introduced a notion of the real or ideal as a presupposition. To read this sentence, one must first accept the statement "As for the *ideal* which we love" as understood. How does a reader do this? He or she must have some prior knowledge of the ideal and must recognize that this knowledge is valorized among individuals with whom the reader identifies. What is this knowledge? Who are these individuals? There are in fact three presuppositions in this sentence, each introduced by a preposition, a coordinating conjunction, or a gerund. By the time the reader gets to the main clause: "in effect we do not love what is outside of ourselves," he or she has already tacitly accepted three statements that do not really assert anything but refer the reader to other unnamed discourses. For Culler, this process of intertextuality does not rely upon the author's intentions or the historical readers' expectations:

In relating [a presuppositional] sentence to other [presuppositions] which form, as it were, its conditions of possibility, one need not enquire whether the speaker of the sentence has previously encountered these other sentences [ . . . ], nor even whether anyone has actually produced these sentences. None of these sentences is a point of origin or moment of authority. They are simply the constituents of a discursive space from which one tries to derive conventions.<sup>13</sup>

The interpretive act by which we make sense of these presuppositions does not simply rely on receiving signs and recognizing their signifieds. Instead, we insert these signifiers into the network of discourses always already present but never fully elaborated during our reading of the text. Intertextual interpretation is therefore the survey of a set of possible meanings that readers attempt to disentangle from a text that is nothing more than fragments from countless other texts knitted together.

Investigating a discursive space can never reach any sort of ultimate mapping. No database can be constructed that would permit researchers to explore every discourse that resonates in a text, especially since cultural, social, and political discourses are not fully transcribed and machine-readable. Nevertheless, databases such as ARTFL enable

us to explore intertextuality in ways that did not exist before computers. In order to pursue the question of realism and idealism in the nineteenth century, one could begin to look for presuppositions about the real and the ideal in the texts from this period. Figure 1 is a kwic (keyword-in-context) listing of words that often function to introduce presuppositions: *quand*, *moment*, *quant*, *lorsque*, etc. The keywords in this list occur in sentences that contain words matching the patterns "idE[aeo].\*" and "rE[ae]l.\*". To make sense out of the seventy-eight hits in this listing, one would have to read at least some of the context for each occurrence and then identify possible discourses running through each sentence containing a presupposition. This step in the analysis is highly interpretive and not amenable to any rigorous methodology. Even if one could accomplish this, the analysis would be incomplete, not only because the database is inherently limited (women writers, non-French writers, and popular works are underrepresented), but also because the discourse of the analysis itself is intertextual. What we today understand by realism and idealism is informed by countless discourses that constitute the history of the disciplines in which we work. We would have to include the history of our own discourse about realism and idealism before we arrive at any conclusion about realist and idealist discourses of the nineteenth century.

Oddly enough, using the ARTFL database is not nearly as problematic if one accepts the contingency of meaning. Those who believe that texts have a unified and transhistorical meaning must look outside the immediate textuality of the database and supplement their research with data from publication histories, social movements, and cultural expressions such as traditions and regionalisms. ARTFL alone cannot reveal the true significance of French literature. However, those who are content with the experience of their own reading will find the database a source of limitless possibilities. The intertextuality that prohibits comprehensive analysis encourages multiple and creative readings.

Anxiety about computers in the humanities may finally have less to do with methodologies and more to do with our expectations for "science." The technology at our disposal leads many individuals to expect that literature will finally be explained scientifically. Unless we make it clear what computers can and cannot do, the myth of science will obscure rather than enhance literary studies.

## References and Notes

1. Mark Olsen, "Texts, Signs and Readers: Quantitative Methods in Socio-Cultural History," in *Proceedings of the Sixth International Conference on Symbolic and Logical Computing* (Madison, S.D.: Dakota State University, 1993), 5-32.

2. Ibid., 10.

3. For more information about ARTFL, including bibliographies, user manuals, and accounts, send e-mail to mark@artfl.uchicago.edu or surface mail to the ARTFL Project, Dept. of Romance Languages and Literatures, University of Chicago, 1050 E. 59th St., Chicago, IL 60637.

4. Jean-Jacques Thomas, "Texts On-Line," *Computers and the Humanities* 27 (1993): 95.

5. Antony Easthope, *Literary into Cultural Studies* (London and New York: Routledge, 1991), 20-21.

6. Roland Barthes, "The Death of the Author," *Image, Music, Text* (New York: Hill and Wang, 1977), 146.

7. Ibid., 147.

8. Naomi Schor, "Idealism in the Novel: Recanonizing Sand," *Yale French Studies* 75 (1988): 61.

9. Ibid., 66.

10. The pattern "idE[aeo].\*" incorporates regular expression

conventions familiar to users of vi, sed, and perl. The E matches the letter e with or without accents. For an explanation of pattern-matching conventions in Philologic 2.0, see the *User's Guide to Philologic 2.0* (Chicago: American and French Research on the Treasury of the French Language, 1993), 20-23.

11. For a brief but adequate discussion of Jauss's *Rezeptionsaesthetik*, see Jonathan Culler, *The Pursuits of Signs* (Ithaca, N.Y.: Cornell Univ. Pr., 1981), 54-58.

12. Maine de Biran, 1824, *Journal T. 2*, ed. H. Gouhier (Neuchâtel, Switzerland: la Baconniere, 1955), 360. (Citation is from the ARTFL database, the emphasis is the author's.) Roughly translated, the quotation reads "As for the ideal which we love, since it is the work of our mind in this system, in loving this ideal, in effect we do not love what is outside of ourselves; but also we do not love nor embrace anything real."

13. Culler, *The Pursuits of Signs*, 117.

# Electronic Texts in the Humanities: A Selected Bibliography

Anita K. Lowry

This is a suggested reference and reading list, whose purpose is to provide librarians with a bibliography of basic sources for understanding how scholars in the humanities use electronic texts and computer-based methods of analysis, for identifying and locating electronic texts and related resources, and for addressing some of the issues involved in the production, distribution, and use of electronic texts.

## Surveys

Unfortunately, there is no basic survey to replace the Hockey and Oakman works from the early 1980s, which now are very dated and are useful primarily as guides to the general categories, vocabulary, and history of humanities and literary computing. The essays contained in *Computers and Written Texts* are much more recent and cover most of the major fields in linguistic and literary computing; its bibliography, though not classified or annotated, is very good.

Butler, Christopher S., ed. *Computers and Written Texts*. Oxford: Blackwell, 1992.

The focus of this volume is on computer applications in linguistics and language studies, except for the following essays: Lou Burnard, "Tools and Techniques for Computer-Assisted Text Processing" (p.1-28); John F. Burrows, "Computers and the Study of Literature" (p.167-204); Wilhelm Ott, "Computers and Textual Editing" (p.205-26). Each essay has a section of suggestions for further reading, and the volume concludes with a twenty-five-page bibliography.

Hockey, Susan. *A Guide to Computer Applications in the Humanities*. Baltimore: Johns Hopkins Univ. Pr., 1980.

Clearly written but dated, this guide covers the techniques for the application of computers to various kinds of linguistic and literary studies.

*Humanities Computing Yearbook*. Oxford: Clarendon Pr. Vol. 1, 1988; Vol. 2, 1989-90.

Although more a directory than a survey (see entry under "Guides and Directories"), it does include a brief state-of-the-art survey at the beginning of each chapter.

Oakman, Robert. *Computers for Literary Research*. Athens, Ga.: Univ. of Georgia Pr., 1984.

This work was originally published in 1980; the bibliography was updated in 1984.

## Guides and Directories

There is no comprehensive source of information about the existence and availability of scholarly electronic texts and text-analysis software in the humanities. The cataloging records of the *Rutgers Inventory of Machine-Readable Texts in the Humanities* (included in the MDF file of the RLIN database [see below]) represent the largest single list of titles; and Columbia University Libraries has recently received a grant to fully catalog a number of the texts and text databases in its Electronic Text Service, which will add several hundred records to the RLIN MDF file and to the OCLC database. The *Humanities Computing Yearbook* and the *Georgetown University Catalogue of Projects in Electronic Text* provide information primarily about electronic text projects and databases, less frequently about individual texts. Other sources of information include announcements and catalogs from publishers like Oxford University Press Electronic Publishing Division, IntelLex Corp., Chadwyck-Healey, Eastgate Systems (hyperfiction), and Intellimation and Voyager (hypertext and hypermedia); announcements and reviews in print and electronic journals and newsletters (including the myriad disciplinary and topical electronic lists and discussion groups); publications of the Center for Electronic Texts in the Humanities (New Brunswick, N.J.) and the Computers in Teaching Initiative Centre for Textual Studies (Oxford University Computing Service); and conferences devoted to computer applications in the humanities, such as the annual joint conference of the Association for Computers and the Humanities and the Association for Literacy and Linguistic Computing.

*Computers in Teaching Initiative Centre for Textual Studies. Resources Guide*. March 1991-. Annual. Oxford: CTI Centre for Textual Studies, Oxford University Computing Services.

This is a very selective annotated guide to commercial and noncommercial electronic texts and text databases, software, and information sources.

*Georgetown University Catalogue of Projects in Electronic Text. Database*. Georgetown: Georgetown Univ., Center for Text & Technology.

An ongoing effort to identify commercial and non-commercial publishers, centers, and research projects around the world producing electronic texts, the database generally does not list individual texts, but gives informa-

Anita K. Lowry is Head, Information Arcade, Main Library, University of Iowa, Iowa City.

tion about the type and extent of work being done along with address and contact information. At present CPET is accessible over the Internet as an interactive Ingres-based system (Telnet guvax3.georgetown.edu, log in as CPET) or as a series of subject- and language-based indexes on Georgetown's gopher (gopher to guvax.georgetown.edu). A major updating and revision of the database is in preparation, and it will soon be available over the Internet as a WAIS-searchable database. For more information contact Paul Mangiafico (pmangiafico@guvax.georgetown.edu or pmangiafico@guvax.bitnet).

Hughes, John J. *Bits, Bytes & Biblical Studies: A Resource Guide for the Use of Computers in Biblical and Classical Studies*. Grand Rapids, Mich.: Academie Bks., 1987.

Dated now, but still useful, *Bits, Bytes & Biblical Studies* provides extensive descriptions and evaluations of software and electronic resources, focusing on those of interest to scholars in biblical, classical, and ancient studies.

*Humanities Computing Yearbook*. Oxford: Clarendon Pr. Vol. 1, 1988; Vol. 2, 1989-90.

Here is the most comprehensive single source of information about databases (textual and otherwise), software, hardware, publishers, organizations and institutions, research projects, and publications relating to all aspects of computing in the humanities disciplines, including history and law.

Oxford Text Archive. *A Shortlist of Machine-Readable Texts Held at Oxford Text Archive*. Latest edition is 1993-94. Oxford: Oxford Text Archive, Oxford Computing Service, 1993.

The catalog lists titles deposited, primarily by scholars, in the Oxford Text Archive; many of the texts are available for distribution at low cost to individuals who sign a written agreement stipulating the conditions of use of the text. The latest edition lists 1,336 titles in twenty-eight languages. A limited number of electronic texts, with TEI markup, are available from the archive via anonymous ftp, as is the *Shortlist* itself and order form; ftp address: ota.ox.ac.uk.

*Research Libraries Information Network (RLIN)*. Database. Mountain View, Calif.: Research Libraries Group. The RLIN MDF file includes MARC records for electronic texts cataloged by member libraries and by the *Rutgers Inventory of Machine-Readable Texts in the Humanities* (see below).

*Rutgers Inventory of Machine-Readable Texts in the Humanities*. Database. This project, begun in 1983 by Marianne Gaunt at

Rutgers University Libraries and now under the aegis of the Center for Electronic Texts in the Humanities, has provided machine-readable cataloging records (available in the RLIN database) for over a thousand electronic texts, including texts in the artfl database and the oxford text archive.

Strangelove, Michael, and Diane Kovacs, comps. Ann Okerson, ed. *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*. 3d ed. Washington, D.C.: Assn. of Research Libraries, 1993.

This directory is an essential source of information about newsletters and lists relating to electronic texts, for example "medtext, [which deals with] medieval texts—philology, codicology, and text analysis" (p. 205). This edition also includes two articles and two annotated bibliographies; the excellent annotated "bibliography of current articles related to electronic journal publications and Publishing" (David F. W. Robison) cites many items relevant to the publication and network distribution of electronic texts.

## Journals, Newsletters, and Electronic Lists/Discussion Groups

Because of their currency, serial publications, both print and electronic, are indispensable to the librarian or scholar wanting to keep up with resources, applications, and activities relating to electronic texts. For more specialized titles devoted to particular humanities disciplines, see the *Humanities Computing Yearbook* and the *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* (listed under "Guides and Directories"), as well as standard bibliographic sources for serials.

*Bits & Bytes Review: Reviews and News of Computer Products and Resources for the Humanities* 1, no. 1 (Oct. 1986- ). Whitefish, Mont.: Bits & Bytes Computer Resources.

This newsletter has a very irregular publication history, but it features detailed and in-depth reviews of software and electronic resources, including electronic texts and text-analysis software.

CETH. Electronic distribution list. List owner: Annelies Hoogcarspel, Center for Electronic Texts in the Humanities, New Brunswick, N.J.

An electronic list established in 1993 to distribute information about the activities of the CETH. To subscribe send an e-mail message containing only the line: SUBSCRIBE CETH firstname lastname to listserv@pucc.bitnet or listserv@pucc.princeton.edu.



*CETH Newsletter* 1, no. 1 (Spring 1993– ). New Brunswick, N.J.: Center for Electronic Texts in the Humanities.

A substantial newsletter that provides excellent overviews of current activities and information sources relating to electronic texts.

*Computer-Assisted Research Forum: A Reader-friendly Bulletin for Academics and Educators in the Humanities* (Fall 1992– ). 3 issues per year. Montreal: Todd Blayone, McGill Univ.

Each issue includes a practical article and several software reviews (comparative reviews are common).

*Computers and Texts*, no. 1 (May 1991– ). 2 issues per year. Oxford: CTI Centre for Textual Studies & Office for Humanities Communication, Oxford Univ. Computing Service.

This newsletter presents articles, reviews, and news notes, with special emphasis on the use of electronic texts and humanities computing applications in teaching.

*Computers and the Humanities* 1 (Sept. 1966– ). Bimonthly. Dordrecht, Neth.: Kluwer.

This is the official journal of the Association for Computers and the Humanities. In addition to scholarly articles on the uses of electronic texts and other computer-based resources and methods for research and teaching in the humanities, it contains good reviews of books, software, and courseware.

*ETEXTCTR*. Electronic discussion list. Moderator: Annelies Hoogcarpsel, Center for Electronic Texts in the Humanities, New Brunswick, N.J.

An electronic list established in 1993 (as a result of the first meeting of the ALA ACRL Discussion Group on Electronic Text Centers) for discussion of issues relating to electronic text centers, including for example, budgets, acquisitions, cataloging, public services, management, training, and staff development. To subscribe send an electronic-mail message containing only the line: SUBSCRIBE ETEXTCTR firstname lastname to listserv@rutvm1.bitnet or listserv@rutvm1.rutgers.edu.

*HUMANIST*. Electronic discussion list. Moderator: Elaine Brennan, Women Writers Project, Brown Univ., Providence, R.I.

Created in 1987, *HUMANIST* is the oldest and most eclectic of the discussion lists related to humanities computing; the list was moderated from 1987 to April 1990 by Willard McCarty, Centre for Computing in the Humanities, University of Toronto, then by Elaine Brennan and Allen Rinear of Brown University. *HUMANIST* also stores and makes available to its subscribers a variety of electronic

information files, many of them relating to electronic texts. To subscribe send an e-mail message containing only the line: SUBSCRIBE HUMANIST firstname lastname to listserv@brownvm.bitnet or listserv@brownvm.brown.edu.

*Literary and Linguistic Computing* 1, no. 1 (1986– ). Quarterly. Oxford: Oxford Univ. Pr.

This journal is an official publication of the Association for Literary and Linguistic Computing. Narrower in scope than *Computers and the Humanities*, it features scholarly articles on the application of computer-based resources and methods in literary and linguistic research, as well as conference announcements and reports and news notes.

*REACH: Research & Educational Applications of Computers in the Humanities* 1, no. 1 (Dec. 1989– ). Quarterly. Santa Barbara, Calif.: Humanities Computing Facility, Univ. of California at Santa Barbara.

Short articles and news notes about activities of the Humanities Computing Facility, applications of electronic texts and humanities computing at UC Santa Barbara, conference notes, and so on are found here. Especially good source for announcements of new electronic journals and lists in the humanities. Also available electronically, by subscription and through the UCSB Humanities Computing Facility gopher.

*SCHOLAR: Natural Language Processing*. Electronic newsletter. Editor: Joseph Raben.

Each electronic issue includes news and notes on books, software, electronic resources, conferences, and other items of interest to scholars in linguistics, literature, history, and lexicography. To subscribe send an electronic-mail message with only the line: SUBSCRIBE SCHOLAR firstname lastname to listserv@cunyvm.bitnet or listserv@cunyvm.cuny.edu.

*Text Technology* 1, no. 1 (Jan. 1991– ). Bimonthly. Madison, S.D.: Dakota State Univ.

Each issue contains three practical how-to articles and reviews on various aspects of "using computers for creating, processing, communicating, and analyzing texts" and a selected annotated bibliography of articles and reviews with a similarly practical focus and emphasis on word processing, computers in composition, etc.

## Books and Articles

This is a highly selective list of works that deal with some aspect of the creation, distribution, or use of electronic texts in the humanities, in particular in literary and

linguistic studies. I have emphasized surveys and anthologies that cover a range of topics or present various viewpoints. Browsing the bibliographies of the works listed here, as well as issues of *Computers and the Humanities* and *Literary and Linguistic Computing* and volumes of *Research in Humanities Computing*, will readily lead the reader to other important titles and to more specialized works.

*The ARTFL Project Newsletter* 8, no. 1 (Winter 1992/93). Chicago: ARTFL Project, Dept. of Romance Languages and Literatures, Univ. of Chicago.

This issue contains six descriptions of how faculty are using a large text database, in this case the ARTFL (American and French Research on the Treasury of the French Language) database, for research and teaching. Newsletter issues may be requested by electronic mail from ARTFL at the address: artfl@artfl.uchicago.edu.

Basch, Reva. "Online Books: Visions, Plans, and Perspectives for Electronic Text." *Online* 15 (July 1991): 13-23.

An introductory survey of the burgeoning developments in electronic text, this article emphasizes resources accessible over the Internet and covers popular and "mass-market" as well as scholarly efforts.

Bolter, Jay David. *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, N.J.: L. Erlbaum, 1991.

A provocative exploration of the fundamental differences between the printed book and electronic text and their implications for reading and understanding texts.

Brockbank, Philip. "Towards a Mobile Text." In *The Theory and Practice of Text-Editing*. Edited by Ian Small and Marcus Walsh, 90-106. Cambridge: Cambridge Univ. Pr., 1992.

This presents speculation and desiderata from a renowned textual editor (general editor of the Cambridge edition of Shakespeare) regarding an interactive optical disk Shakespeare Variorum.

Burnard, Lou, C. M. Sperberg-McQueen, and Susan M. Hockey. *An Introduction to TEI Tagging* [Chicago: Text Encoding Initiative], July 28, 1993.

This is a short introduction, with examples, to the major features of TEI markup for electronic texts. Available as document number TEI U 1 from the TEI listserv (see *Text Encoding Initiative* entry, below).

CCH *Working Papers* 1 (1991- ). Occasional. Toronto: Centre for Computing in the Humanities, University of Toronto.

This excellent series of small volumes covers the

nature and use of electronic texts and text-analysis software for literary and linguistic studies.

- Vol. 1: *A TACT Exemplar* (1991)
- Vol. 2: *Historical Dictionary Databases* (1992)
- Vol. 3: *Computer-based Chaucer Studies* (1993)

Chernaik, Warren, Caroline Davis, and Marilyn Deegan, eds. *The Politics of the Electronic Text*. Oxford: Office for Humanities Communication, Oxford Univ. Computing Services, 1993.

Essays, both practical and theoretical, discuss various topics relating to access and control of "the electronic text [which] is both more accessible and more provisional and impermanent than the conventional printed text" (p. 6). Several essays are concerned with the scholarly editing process in the electronic age.

"Computers and the Humanities." Special issue of *Bulletin of the John Rylands University Library of Manchester* 74, no. 3 (Autumn 1992).

See especially the following articles: David L. Mealand, "On Finding Fresh Evidence in Old Texts: Reflections on Results in Computer-Assisted Biblical Research" (p. 67-87); Kathryn Sutherland, "Challenging Assumptions: Women Writers, the Literary Canon and New Technology" (p. 109-20); Allen Renear, "Representing Text on the Computer: Lessons For and From Philosophy" (p. 221-48).

Coombs, James H., Allen H. Renear, and Steven J. DeRose. "Markup Systems and the Future of Scholarly Text Processing." *Communications of the ACM* 30 (Nov. 1987): 933-47.

An influential analysis of text-markup systems and argument for the use of descriptive markup in machine-readable texts. Reprinted in George Landow and Paul Delany, eds. *The Digital Word: Text-based Computing in the Humanities*, 85-118. Cambridge: MIT Pr., 1993.

Crane, Gregory. "'Hypermedia' and Scholarly Publishing." *Scholarly Publishing* 21, no. 3 (Apr. 1990): 131-55.

A cogent discussion of the role of scholarly hypermedia databases in revitalizing the nature and use of the "monograph," "book," and "textbook," this illustrates the argument with examples drawn from potential uses of *Perseus: Interactive Sources and Studies on Ancient Greece*.

Daly, James, ed. *Workshop on Electronic Texts: Proceedings*. Washington, D.C.: Library of Congress, 1992.

This 118-page work summarizes presentations and discussion at an invitation-only workshop on electronic texts; in spite of the necessarily "episodic" and fragmentary nature of this attempt to capture the lively give and

take of the workshop, it contains useful information on a variety of technological and access issues related to the design, production, preservation, and use of electronic text and image databases.

Delany, Paul, and George P. Landow, eds. *Hypermedia and Literary Studies*. Cambridge: MIT Pr., 1991.

Essays by various writers discuss the theory and practice of scholarly hypertext, including detailed descriptions of specific hypertext and hypermedia projects in classics, biblical studies, literature, and theater.

"Economic Models for Networked Information." Special issue of *Serials Review* 18, nos. 1/2 (1992).

Though they do not deal specifically with electronic primary source texts in the humanities, the twenty-five articles in this issue make important contributions to the ongoing discussion of the difficult economic and intellectual property issues involved in network distribution of and access to information.

Ginsburg, Jane C. "Copyright Without Walls? Speculations on Literary Property in the Library of the Future." *Representations* 42 (Spring 1993): 53-73.

This article presents a clear, concise discussion of copyright issues raised by the digitization of library holdings.

Goldfarb, Charles. *The SGML Handbook*. Oxford: Oxford Univ. Pr., 1991.

The handbook is a comprehensive guide to the Standard Generalized Markup Language (SGML) (ISO 8879).

Hockey, Susan. "Developing Access to Electronic Texts in the Humanities." *Computers in Libraries* 13, no. 2 (Feb. 1993): 41-43.

Explained here are the goals and current activities of the Center for Electronic Texts in the Humanities, including cataloging electronic texts, supporting the Text Encoding Initiative, collecting and disseminating texts, and educational programs and support services.

Kenny, Anthony. *The Computation of Style*. New York: Pergamon, 1982.

This work is generally considered the best introduction to statistical and computational methods for humanists.

Landow, George P. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: Johns Hopkins Univ. Pr., 1992.

The relationship between hypertext and contemporary theory is examined, and the impact of literary hypertext on the idea and nature of the text and the author,

on narrative, on education, and on the politics of information ("Who Controls the Text?") is considered.

Lowry, Anita. "Machine-Readable Texts in the Academic Library: The Electronic Text Service at Columbia University." In *Computer Files and the Research Library*, 15-23. Edited by Constance C. Gould. Mountain View, Calif.: Research Libraries Group, 1990.

The rationale for and description of one of the earliest electronic text facilities in a library emphasizes public services issues.

Lynch, Clifford A. *Accessibility and Integrity of Networked Information Collections*. Background Paper BP-TCT-109. Washington, D.C.: Office of Technology Assessment, U.S. Congress, 1993.

Although it does not deal with electronic texts per se, this report provides a wide-ranging survey of the complex issues raised by networked access to information.

McCarty, Willard. "A Potency of Life: Scholarship in an Electronic Age." In *If We Build It: Scholarly Communications and Networking Technologies: Proceedings of the North American Serials Interest Group, Inc. 7th Annual Conference, June 18-21, 1992*. Edited by Suzanne McMahon, Miriam Palm, and Pam Dunn. Also published in *The Serials Librarian* 23, nos. 3/4 (1993): 79-97.

This eloquently presented view discusses how the "dynamic, changeable nature of electronic data" (p. 81) can influence communication and, through electronic texts and text-analysis software, research in the humanities.

Neuman, Michael. "The Very Pulse of the Machine: Three Trends Toward Improvement in Electronic Versions of Humanities Texts." *Computers and the Humanities* 25 (1991): 363-75.

An overview is given of trends in scholarly and commercial electronic text production and publication.

Olsen, Mark. "The Language of Enlightened Politics: The 'Société de 1789' in the French Revolution." *Computers and the Humanities* 23, nos. 4/5 (Oct. 1989): 357-64.

An analysis is presented on how a large electronic text database and computer-based methodologies can be used for studies in the history of ideas.

Olsen, Mark. "Signs, Symbols and Discourses: A New Direction for Computer-Aided Literature Studies." With responses. *Computers and the Humanities* (forthcoming).

Distributed and discussed widely over the networks and presented at a session of the Modern Language Association annual meeting, this paper has engendered a stimulating debate about the theoretical and practical foundations of computer-aided literary textual analysis. Olsen proposes a new paradigm for such studies, based not on the close statistical analysis of individual works but on the use of large text databases to investigate the literary and historical context of works and ideas (to greatly condense and simplify his argument). The publication of this paper and the responses to it promise to provide insight into contemporary issues in literary computing.

Potter, Rosanne G., ed. *Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric*. Philadelphia: Univ. of Pennsylvania Pr., 1989.

A classic collection, these essays discuss or demonstrate the application of computational methods in literary criticism.

Price-Wilkin, John. "Text Files in Libraries: Present Foundations and Future Directions." *Library Hi Tech* 9, no. 3 (1991): 7-44.

A review and update of the survey of electronic texts in RLG libraries (see below), the article gives an extensive description of "UMLibText . . . the University of Michigan Library's system for wide area network-based access to software and texts needed for textual analysis" (p. 15).

Sidebars are included on the ARTFL database, the Oxford Text Archive, the IntelLex Corp. (a publisher of scholarly electronic texts in philosophy), the Open Text Corporation (publisher of Pat & Lector text-analysis software), the Text Encoding Initiative, the electronic edition of the *Oxford English Dictionary* (second edition), and the Center for Electronic Texts in the Humanities.

Price-Wilkin, John. "Text Files in RLG Academic Libraries: A Survey of Support and Activities." *Journal of Academic Librarianship* 17, no. 1 (Mar. 1991): 19-25.

Proud, Judith K. *The Oxford Text Archive*. British Library Research & Development Report 5985. London: British Library, 1989.

This commissioned study of the current state of the Oxford Text Archive and its holdings makes recommendations for future developments.

*Research in Humanities Computing* 1, (1991- ). Annual. Oxford: Oxford Univ. Pr.

This series publishes selected papers from the annual joint conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic

Computing, beginning with the 1989 conference held in Toronto.

Robinson, Peter. *The Digitization of Primary Textual Sources*. Oxford: Office for Humanities Communication, Oxford Univ. Computing Service, 1993.

This detailed introduction to the practical and technical aspects of creating facsimile digital images of books and manuscripts provides descriptions of selected digitization projects, including the Cornell/Xerox "brittle books" joint study, the program for the new Bibliothèque de France, the Wittgenstein archives, the Archivo General de Indias project (Seville, Spain), the Canterbury Tales project, and others.

Robinson, Peter. *The Transcription of Primary Textual Sources Using SGML*. Oxford: Office for Humanities Communication, Oxford Univ. Computing Service (forthcoming, 1993).

Seaman, David. "The Electronic Text Center On-line Archive of Electronic Texts." In *Scholarly Publishing on the Electronic Networks. The New Generation: Visions and Opportunities in Not-for-Profit Publishing*. Proceedings of the Second Symposium, December 5-8, 1992, edited by Ann Okerson, 101-06. Washington, D.C.: Office of Scientific and Academic Publishing, Association of Research Libraries, 1993.

A short description tells of the facilities and services of the Electronic Text Center in the Alderman Library at the University of Virginia and its networked system for access to and use of electronic texts.

Sperberg-McQueen, C. M. "Text in the Electronic Age: Textual Study and Textual Encoding, with Examples from Medieval Texts." *Literary and Linguistic Computing* 6 (1991): 34-46.

This article "describes, with examples, the theory implicit in the work of the Text Encoding Initiative (TEI), a project to develop guidelines for the encoding of machine-readable texts" (p. 34) and presents a clear and compelling case for the assertion that "anyone who wishes to create serious electronic texts must deal seriously with the characteristics of markup" (p. 35).

Sperberg-McQueen, C. M., and Lou Burnard, eds. *Text Encoding Initiative. Guidelines for Electronic Text Encoding and Interchange*. Rev. version. Chicago: Text Encoding Initiative (forthcoming).

Expected to be more than a thousand pages when published, this "final" version (also known as TEI P3) of the TEI guidelines (an implementation of SGML) will pro-

---

vide detailed specifications for markup for all kinds of scholarly electronic texts.

Preliminary versions of the guidelines, by section, as well as numerous reports and supporting documents of the TEI are available from the TEI listserv at the University of Illinois at Chicago Circle; for a list of documents send an electronic mail message to listserv@uicvm (Bitnet) or listserv@uicvm.uic.edu (Internet) with the command GET TEI-L FILELIST.

Stover, Mark, ed. "Electronic Information for the Humanities," *Library Trends* 40, no. 4 (Spring 1992).

See especially the following articles: Anita Lowry, "Electronic Texts in English and American Literature" (p.704-23); Edward Shreeves, "Between the Visionaries and the Luddites: Collection Development and Electronic Resources in the Humanities" (p.579-95); Mark Stover, "Religious Studies and Electronic Information: A Librarian's Perspective" (p.687-703); Erwin K. Welsch, "Hypertext, Hypermedia, and the Humanities" (p.614-46).

Van Herwijnen, Eric. *SGML Tutorial*. Version 2.0. Computer file. Providence, R.I.: Electronic Book Technologies, 1993.

This electronic edition of van Herwijnen's book *Practical SGML* (Dordrecht, Neth.: Kluwer, 1990) contains extensive beginning and advanced interactive tutorials and exercises to teach SGML and uses DynaText software to manage, browse, and search the text, thus demonstrating the features of one of the most widely known programs available for SGML marked-up text.

Virbel, Jacques, "Reading and Managing Texts on the Bibliothèque de France Station." In *The Digital Word: Text-Based Computing in the Humanities*, edited by George Landow and Paul Delany, 31-51. Cambridge, Mass.: MIT Pr., 1993.

An interesting description of the functionality envisioned for "a specialized [microcomputer-based] reading station for accessing and manipulating documents and for linking the user to the whole collection of resources (materials, software, and databases)" (p. 32) to be developed for the new Bibliothèque de France (to open in 1995); the project to put approximately 350,000 works into digital form for the library is briefly described in Robinson, *The Digitization of Primary Textual Sources* (see above).

# Authority Control

Consistent and current authority headings are the key to the effectiveness of any library catalog. The *WLN MARC Record Service (MARS)*<sup>TM</sup> compares your headings to LC Name and Subject, LC Children's, NLM medical subject and/or WLN authority files; verifies or upgrades all headings to current, authorized forms; and supplies a complete and accurate set of matching USMARC authority records including notes and cross references. At WLN, we offer personalized service throughout your project, strict adherence to project schedules, comprehensive bibliographic record and authority heading upgrading programs, customized programming based on your specifications, auxiliary authority files for improved match rates, and three levels of manual review. Most of all, we offer authority control tailored to your needs at a price you can afford. Let the *MARS* team be your extra staff member, invisible yet indispensable.

*WLN MARC Record Service: Authority Control . Database Preparation . Customized Programming . Smart Barcode Labels . Bibliographic Record Upgrading . Manual Review . CD-ROM Public Access Catalogs . Ongoing Authority Updating/Database Updating, available via FTP.*

Yes, I want to find out more about authority control and how the  
*WLN MARC Record Service (MARS)* can help my library.

Name \_\_\_\_\_ Title \_\_\_\_\_

Institution \_\_\_\_\_

Address \_\_\_\_\_

Telephone \_\_\_\_\_ Best time to be reached \_\_\_\_\_ a.m. \_\_\_\_\_ p.m.

Type of project:  Authority control  Holdings/item conversion  Smart barcodes  Other \_\_\_\_\_

If authority control:  LC Name/Subject Authority File  LC Children's  NLM MeSH

WLN Authority File  Updating Service  Other \_\_\_\_\_

Size of database \_\_\_\_\_ Local system vendor \_\_\_\_\_ Source of MARC records \_\_\_\_\_

Projected start date \_\_\_\_\_ Completion date \_\_\_\_\_

Comments \_\_\_\_\_

Signature \_\_\_\_\_



TURNING INFORMATION INTO KNOWLEDGE

WLN . P.O. Box 3888 . Lacey, WA 98503-0888 . 1-800-DIALWLN . (206) 923-4000 . FAX (206) 923-4009

# SPECIAL SECTION

## Future Possibilities in Information Technology and Access

### Introduction: Science Fiction— A Mainstream Force

Milton Wolf,  
Special Section Editor

Two well-known writers, Connie Willis and David Brin, spoke to an overflow crowd at the ALA Annual Conference in New Orleans. Sponsored by the Imagineering Interest Group of LITA as part of an ongoing series of talks that engage highly creative people to imagine the future, particularly as it pertains to the information world, this informal forum has proved exceptionally enlightening and has offered a stellar group of prescient commentators: Frederik Pohl, Hans Moravec, Bruce Sterling, Pat Cadigan, Tom Maddox, Steve Cisler, David Porush, Elizabeth Moon, to mention only a few.

While the majority of us would admit that we are not as scientifically

literate as we should be, few of us would as freely concede that we are also ignorant of the advances in literature. Most of us confuse "sci-fi," which really refers to movies (e.g., *Star Wars*, *2001*, *RoboCop*, and so on) with the 1920s and 1930s science fiction pulp magazines and their garish covers depicting bug-eyed monsters and scantily clad women being abducted by aliens. Well, Virginia, times have changed and such clichéd thinking needs to be discarded.

Ursula K. LeGuin has remarked that between 1960 and 1990 the genre of science fiction (SF) came of age, "moving from modernism to the post-modern." The most recent edition (1988) of the *Columbia Literary History of the United States*, under the section "The Fictions of the Present," lists SF as "arguably the most significant body of work in contemporary fiction," citing SF as "a major literary genre" and one of the "most significant new directions in recent American fiction."

that made those velociraptors possible is going to make it possible in very short order to do digitized stunts and sets. And even stars. It will be possible to remake *Casablanca* with Ronald Reagan, who was originally supposed to star in the movie. It will be possible to take the smoking out of *Now, Voyager* and change the ending of *Gone with the Wind*.

I was expressing the opinion that these possibilities did not seem to me to be an unmixed blessing. And one of the panelists said, "You have to keep in mind, Connie's a technophobe."

My first response was, "I am *not*! I'm a science fiction writer. That's what we do."

It is what we do. The job of a science fiction writer is to look for the unforeseen consequences, the problems and kinks and side effects that

The dialogue between science and fiction is a cross-fertilization that has perforce moved SF into the cultural mainstream. What American jazz is to world music, SF is to world literature. It is a literature of anticipatory processes, positing a variety of future worlds, and since many of society's directions will be selected from the images and metaphors created by its imagination, a poverty in this area limits its choices.

Unfortunately, ALA events number in the hundreds—especially on Sunday afternoons when the Imagineering Group meets—so the following modified transcripts of the talks of Connie Willis and David Brin are being published for those of you who missed the insightful remarks by these two creative observers of the information revolution. (Connie Willis recently won the prestigious Nebula Award for *Doomsday Book*, and David Brin's *Earth* is a fictional tour de force of the possible social ramifications of information technology.)

nobody has really thought about. Ed Bryant said it best. He said that if this were 1890, it would take an inventor to predict the automobile, and it would take a real visionary to predict highways and gas stations. But it would take a science fiction writer to predict the traffic jam.

Science fiction writers have been predicting traffic jams in the information revolution for years. In "And Now the News," Theodore Sturgeon worried about the terrible, unforeseen effects that can result from too much media exposure and too much bad news. It was written in the 1950s, but it reads as if it were composed from yesterday's headlines.

Gordon Dickson fretted and stewed about the dangers of computers in the hilarious "Computers Don't Argue," in which a man tries to return

### **Jurassic Park and Al Jolson: Thinking about the Information Revolution**

Connie Willis

A couple of weeks ago I was on a panel about the information revolution, and we were talking about electronic publishing and computer graphics. And *Jurassic Park*. I had a lot to say.

I have been working on a short novel, called *Still Frame*, about the future of the movies. Hollywood is in the middle of a revolution, whether it realizes it or not. The computer graphics

a murder mystery to the book club he didn't order it from and ends up being executed for the murder in the title.

And, of course, there is that star-class worrier, George Orwell and his 1984. Remember how we all congratulated ourselves in 1984 that here it was and it hadn't happened after all? And ten minutes later, here it is: Newspeak and the omnipresent TV screens and electronic surveillance and manipulation of the news; Big Brother in all his glory.

Science fiction writers have been expressing reservations about the information revolution for years, I thought. It is their job. It is my job. I am not a technophobe. I am not afraid of technology.

And then I thought, Well, maybe I am afraid of technology a little. And then I thought, Maybe it is a good idea to be afraid of technology, especially the revolutionary kind.

If you have seen *Jurassic Park* you will recall the scene in which they first see a dinosaur. It is a brachiosaurus, and, all excited, they go rushing over to where it is grazing. Now, it is true that the brachiosaurus is a herbivore. But just because it won't eat you doesn't mean that it won't swing that neck around and knock you out of the picture. Just because it doesn't have evil intentions doesn't mean that it won't step on you. And so can the information revolution.

This doesn't mean that I am against having an information revolution. We are already having it whether we want it or not and there is nothing we can do about it. In *Jurassic Park* there is a scene in which the chaos scientist is telling people that they shouldn't have [cloned the dinosaurs]. He says something to the effect that you only thought about whether you could do it, you didn't think about whether you should do it.

This is, of course, [one of] the old Frankenstein "there are some things man was not meant to know" speeches that have been made in the movies for years. Personally, when the

chaos scientist made the speech, I wished that the tyrannosaurus had reached down and bitten off his head, because when you are watching the movie, half of you is thinking, "Bad idea to bring the dinosaurs back! Bad idea," while the other half is thinking, "I would kill to see a dinosaur!" That is pretty much how I feel about the information revolution. I think that it is going to be absolutely wonderful. I just don't want to get too close and get stepped on. I just think we should think about what traffic jams might lie ahead.

This brings me to Al Jolson and the talkies. One of the things that makes my fellow writers think of me as a technophobe is that I am always talking about the past. When I need to extrapolate the future, I always try to see how things were in similar situations in the past. When I was surveying a planet for my new novella, *Uncharted Territory*, I went back and read about Spekes and Burton in Africa and Jim Bridger and Kit Carson and the Yellowstone expedition.

When I wanted to create a modern-day epidemic in my novel *Doomsday Book*, I went back and looked at the Spanish flu epidemic of 1918, which was the first pandemic and killed more people than World War I. And when, in the same book, I wanted to show the end of the world, I went back to the first time it ended, to the Black Death in the 1300s, which swept down like an avenging angel and wiped out one-third to one-half of Europe in less than five years.

When I wanted to talk about an information revolution, I went back and looked at the first time it happened to Hollywood, in 1927. When Al Jolson said, right up there on the screen, in words you could hear, "You ain't seen nothin' yet!" he set off changes just as revolutionary as the ones we are looking at right now.

Everything suddenly turned topsy-turvy as every studio in Hollywood raced to get on the bandwagon. The studios shut down so they could

put in new sound equipment. Silent movies were reshot and given sound tracks. Movie stars signed up to take elocution lessons (you've seen this in *Singin' in the Rain*), and everybody had to learn how to talk into bushes and flower arrangements.

Old jobs, like the piano player who had played the mood music on the silent set, suddenly became obsolete, and new jobs, like sound man and electrician, sprang into being. Some people predicted that talkies were nothing but a passing fad, and some people claimed they would create a wonderful utopia, educate the masses, and cure the common cold. And everybody converted to the new system, which was horrible. It was hopelessly complicated to use, it took twice as long as the old system to do anything with it, half the time it didn't work, and the whole system was obsolete even before it was installed.

Does any of this sound familiar? Does it sound like the information revolution we are in right now?

What can we learn from this other revolution? We can learn not to throw the baby out with the bathwater. The first thing Hollywood did after Al Jolson spoke was to dump all of their silent stars and race off to New York to sign up Broadway actors. It didn't work. The movies and theater are two different things, and the Broadway actors, with a couple of exceptions, looked stagy and artificial on the screen. They were also not user-friendly. People went to the movies to see familiar faces, and after a couple of years the studios came to their senses and let the silent stars talk, which they could have done in the first place.

We can also learn to think first and then act (never a common human trait). We can try to imagine the consequences before we go plunging into the middle of this information revolution. The movie studios had already chained themselves to "All Talking, All Singing" before it occurred to them that they had destroyed one of their biggest sources of revenue, their for-



eign markets that understood the silent films perfectly but didn't speak English. Hollywood had to regroup frantically and reinstate foreign-language title cards.

It also didn't occur to Hollywood that the talkies might talk but that was about all they did. They were so eager to get one hundred percent sound in their movies, they ended up with the whole cast grouped around the coffee table shouting their lines into the vase with the microphone hidden in it. The outside setting disappeared, because there was too much noise, as did the train wreck, the dramatic chase on horseback, and the scene on the ice floe, all of which the public went to see.

And it did not occur to Hollywood what the biggest impact on them might be. John Gilbert was the biggest romantic star of the silents—even bigger than Rudolph Valentino—and most people remember him as the actor who couldn't make the jump to talkies because of his voice. According

to the legend, he opened his mouth for his first big talkie love scene, and a thin, high voice came out and ended his career.

It is true, his voice was rather reedy (so is Tom Selleck's, who doesn't have any problems being a romantic lead), and, it is true, audiences did laugh, but not because of his voice. They laughed because nobody in Hollywood had thought of the consequences of what they were doing. Audiences had never heard a love scene before. It wasn't just John Gilbert's first talkie love scene, it was the first for moviegoers, too, and they did not know how to handle it. The silent-movie-going audience had watched the screen and made up the words, the voices, and much of the plot was in their own minds. It was a very private affair. Hearing a love scene was like eavesdropping on a couple in a bedroom. It was embarrassing, and when people are embarrassed they laugh.

It had also not occurred to anyone

in Hollywood that now that the movies had dialogue, they needed scripts. The silents had not had scripts. The actors had made up the dialogue as they went along. For his first big love scene the only lines John Gilbert had to work with were "I love you, I love you, I love you." You would have laughed, too. Unfortunately, it wasn't funny. As a result of that love scene, John Gilbert's career self-destructed, and so did he. An unfortunate side effect of revolutions is that they usually take a heavy human toll.

Thinking about those things in advance could have saved John Gilbert's career and a lot of grief. What was needed was a little forethought. And a healthy dose of technophobia.

What are my predictions for the future of the information revolution? Simple. First, those who don't understand the past are condemned to repeat it. Second, revolutions are a lot like brachiosaurs. And third, "You ain't heard nothin' yet."

---

## The Good and the Bad: Outlines of Tomorrow

David Brin

Recently, I read a newspaper report of great interest to a new parent like me. A firm known as California Technology Enterprise Corporation is marketing an innovative gizmo—a spinoff of high-tech defense hardware—which they call "Beeper Kid," to help keep track of your toddler at the shopping mall or county fair. It lets out a little shriek whenever your three-year-old strays more than ten feet away from the home unit, strapped on your wrist.

Next, of course, will come a little direction finder, to show in which di-

rection the kid has dashed off so recklessly. Great for tracking that wandering child, or finding those missing car keys . . . or for planting under your husband's car, to find out where he *really* spends Tuesday evenings, when he's supposed to be playing poker with the boys.

Isn't that the way of modern technology? All recent generations of humanity have had to deal with this vexing combination. Shiny new gadgets and services often work even better than promised, but they come inevitably laced with countless unexpected consequences; all of which makes life ever more complex, nerve-racking, and dangerous.

Think about the wonders on sale right now at your local Radio Shack. Any of you can saunter in and buy a toy car for twenty bucks, containing a remote control and four or five wire-

less-actuated motors—exactly the kind of stuff Q was always so proud of supplying to secret agent James Bond—only now mass-produced so that ten million home hobbyists can have some cheap, harmless fun. Meanwhile, thousands of would-be inventors have access to inexpensive parts they can experiment with to their heart's content, coming up with countless new ways to do things, some of which may even make life better.

And, oh, yes. Out of millions of customers, and thousands of experimenters, a few dozen of those carrying toy cars out of the corner Tandy Store will be psychotics, having just purchased ideal tools to help manufacture remote-controlled terrorist bombs.

This odd mixture of fulfilled dreams and unforeseen side effects has

been going on for a long time. Mind you, one of the most common misapprehensions about science fiction writers is that we *try* to predict the future. In fact, we only do a little better than stock market analysts and weathermen. Yes, we've had some famous hits—such as air and space travel, nuclear weapons, and eco-devastation. On the other hand, no author before 1979 came even close to predicting something as fundamental to our modern world as the home computer.

No writer did so because everyone at the time was distracted by a red herring. Even the most imaginative authors were suckered into writing novels, stories, and films about *gigantic* computers, occupying whole city blocks, dominating metropolitan skylines. They were all misled by the "obvious."

Why? First off, back in the seventies it certainly looked as if massive central processors would be the way of the future. That's what nearly all the futurists and computer mavens were predicting. Even more important, it suggested two obvious and irresistible possible story lines.

Plot A: The big machine is evil. It takes over and tries to kill everybody. The hero of your story fights it.

Plot B: The computer is good. But it makes a frail, easy target, and your hero has to protect it against some nefarious scheme or saboteur.

Nobody predicted that the home computer would displace the megamachine and go on to replace the rifle over the fireplace as freedom's great emancipator, liberating common citizens as no other technology has since the invention of the plow. (The only writer who came close was John Brunner, in his wonderful novel *Stand on Zanzibar*. Way back in 1967, he had one of his characters say, "Do you know that if we weren't spending so much on arms, every home in America could have a Shalmanisar-level computer?" Brunner just missed. He saw the possibility,

but could not quite bring himself to believe we would actually do it.)

So even science fiction writers must approach prognostication with a sense of humility. Should anyone ever set up a neutral, unbiased "predictions registry," as I depict in my recent novel *Earth*, an organization dedicated to logging, scoring, and/or debunking all claims of prophecy, the grades handed out to SF authors would be above average (certainly far better than so-called psychics), but not so good that any of us ever retired on winnings from the racetrack.

Nevertheless, it's safe to say that many more wonders lie on the horizon that we'll use for well and ill . . . both to improve our culture and at the same time make things even more unmanageable than ever.

For instance, imagine the bright future of *chemical transducers*, sophisticated relatives of your home's smoke detector, which will stand guard to alert us of too much chlorine in the pool or marijuana on the breath of a bus driver. Government and industry are already using early, expensive versions, such as battlefield transducers, which watch over our troops to warn of invisible poison gas. Next we'll see them at border checkpoints, augmenting old Rover, the sniffer dog. Eventually chem-sniffs will show up at K-Mart, to be purchased by home owners worried about ozone, radon, or lead in the home. Or by neighborhood watch committees, keeping an eye on local hazards to public safety. Or by parents, bugging their kids' rooms for traces of cocaine.

You think things are bad for smokers now? Just wait until your coworkers replace those plaques on their desks, saying "Smoke-Free Zone"—with laser alarms that will zap if you had a cigar the night before.

The funny thing about the picture I'm painting is that it can be seen as either good news or bad. In the last example, I guess it depends on whether you smoke or consider tobacco a noxious weed.

In fact, whether we approve or not hardly matters. These things will happen. All we can do is try to be prepared.

Nevertheless, we will be urged to vote in all sorts of new laws to regulate these brave new technologies. Throughout the last five hundred years, there has been fierce tension between two contrary penchants in Western civilization. The first contends that free individuals will make mistakes, but can generally be trusted with new technologies, with new ideas and new freedoms.

The other, puritanical voice worries that human beings are inherently corruptible, that freedom will be abused, and that ideas themselves can often be perilous. Potentially destructive technologies must be kept out of the hands of common folk at all costs, for their own good.

In this perennial debate, it does no good to posture, as do self-righteous purists on both sides. Both parties have plenty of anecdotes on hand, to illustrate and support their view of human nature, because that nature spans a wide spectrum and defies any simpleminded effort to model it.

On the whole, we in this culture tend to lean in the direction of trust, a decision I support wholeheartedly, since the result is a lively, exuberant era, whose wild brilliance merits putting up with a certain amount of chaotic excess. But then I am a product of generations of propaganda extolling freedom as a good thing, as the *best* thing, worth any price to win and hold.

Now a contrarian historian might point out that nearly every other human civilization took an opposing view, that liberty is a dangerous, potentially fatal social illness. From tribes to kingdoms to empires, almost none ever preached as we do, in contemporary films and situation comedies, that individual eccentricity, diversity, and license are sacred traits, to be extolled right up there with (and sometimes above) patriotism and motherhood.

Are we right? Was every other human society wrong? It's a question that is difficult to answer.

Take, for example, the checkered story of a publication entitled *The Anarchist's Cookbook*, which contained recipes for making bombs out of household detergent and other items available off the shelf in any supermarket. When some police functionaries tried to have the book suppressed, the Supreme Court supported freedom of speech. But I wonder, might it even have mattered if would-be censors succeeded in banning it? Here we have a society that sells millions of barrels of liquid high explosives on street corners to all comers . . . which they proceed to pour into vehicles in order to send tons of steel hurtling faster than a cheetah, faster than most *birds*, while the drivers chew gum, fiddle with the radio dial, eat hamburgers, steer with one finger, and go whizzing through tight turns, routinely missing other cars by inches. Who, one hundred years ago, would have predicted that the *most* skilled human being, let alone the rest of us, would be capable of such feats? Who could have predicted we would *want* to?

Such a society simply has to rely on the general good sense of common citizens. So far, at least regarding the public's right to drive, that trust seems not to have been badly placed.

So far.

And yet one can imagine scenarios in which a technology might *have* to be fiercely squelched. Say it becomes feasible, someday, for any family to make a nuclear bomb in their attic. Or, as in my novel *Earth*, to make a mini-black hole in the basement. If such powers really can be Radio Shack'd (now there's a verb for you), then no law, no matter how draconian, would suffice to protect us. If the time ever comes when one lone fool can bring down everything millions of decent people have built, then we finally have the answer to why we've never been contacted by

alien intelligent life. Every other sentient race has wiped itself out long before getting off its home planet.

### The Chemical Pandora's Box

Barring such extreme and catastrophic examples, what sorts of new opportunities might lie on the horizon to tempt us, thrill us, and make ever-greater demands on our good judgment?

Some trends will be unstoppable. For instance, two new technologies will soon make today's quixotic drug laws even more pathetically impossible to enforce than they already are.

Advances in automatic chemical synthesis will in time put machines on the market capable of manufacturing in a bottle almost any known bioactive substance. These will first appear in university labs where, by day, miracles of medicine will be invented to save countless lives. And, late at night, there will also pour from these labs tons of bootleg RU486, the morning-after abortion pill, if that much-coveted drug is not legalized.

A little later, the next generation of synthesis machines will sit in your local pharmacy, so that any rare or special prescription can be filled while-u-wait, without having to send away to distant factories. Then as equipment prices drop with mass production, these devices will appear in local high-school chem labs. (Want to keep them out? And let kids in Japan and China get even *further* ahead of our students? No way!)

After that, can home units be very far behind?

Wonders will spring forth from thousands of basement laboratories. A new realm of convenience and creative living through chemistry . . . as well as a flood of cheap mind-benders, illicitly tailored by certifiable idiots bent, as usual, on spoiling a good thing.

Imagine the social angst! The fervent calls for legislation! But this

phase will soon be superseded, in turn, by yet another. An age when each of us, for better or for worse, shall have the sovereignty to use or abuse chemical marvels, without any need for outside machinery or expertise at all.

How will this be? It is inevitable, as we learn more about the functioning of the finest chemical-synthesis factory of them all—our own living bodies. In particular, we are starting to understand a thing or two about the workings of the human brain, and it is becoming clear that everything we love, every pleasure we partake in, from the most sublime music, to athletics, to sex, to affection and parenthood, appear to be mediated by a swarm of self-released psychoactive compounds.

Now some will be horrified by this revelation. The same folks who decry that the beauty of a rainbow is diminished if we penetrate its secrets to discover that it consists of ten trillion floating watery lenses, all brilliantly refracting, in perfect synchrony, rays from a stellar fusion pile, burning with ancient, furious constancy, millions of miles away.

Indeed, there is no shame in knowing about the psychoactive chemicals released by our brains, the endorphins and enkephalins, for which opiates like heroin are mere shabby substitutes. For instance, my little eighteen-month-old son has something my wife and I call the devil-boy grin. Does he know, instinctively, that his smile, all babies' smiles, will trigger a surge-release of endorphins inside the brains of 90 percent of women and 60 percent of men? All I know is that he saves his very best beam for when he's about to do something he knows is forbidden, and he knows I'm watching. Then in go the dimples and out come the teeth. Those eyes squint and come alight.

Zap! I can *feel* the adoration hormones secreting, and my mind feels like it's about to melt. I laugh. He laughs. We both break down laugh-

ing. Ah, the tribulations of parenthood.

There's no disgrace in learning that chemistry plays a role in life's magic. But what does it imply about our present-day drug laws? What serves wholesome joy can also be abused. It is a simple fact that people seek pleasure, and those who cannot find it in healthy ways have always turned to look for it elsewhere.

Consider the new techniques becoming available through electronic *biofeedback*. Already computer programs are coming out that use visual and aural repetition techniques to help train you to slip into the so-called alpha state almost at will. For most, this will be just another useful tool, such as when you sit down in the dentist's chair and wish to "be somewhere else for a while." On the other hand, any person inclined to misuse these programs might, within a week or so, learn to emulate what Eastern gurus may have been doing for the last twenty-five hundred years, that is, release a flood of their own endorphins, putting themselves into an incognizant stupor in a matter of minutes.

Now today, when a fellow is seen sitting in a park with a line of drool trailing slowly out the corner of his mouth, there's a good chance he's high on something illegal. But say it's ten years from now, and he has no detectable unnatural drug in his veins, just floods of his own, natural secretions. What's a cop to do? Roust him and demand in a loud voice, "Hey buddy! Show me your Zen license"?

In one of my novels, I coin the word "dazer" to describe the next wave of addicts, the self-doping kind, who will need to commit no crimes to support their habit and who will send no money to Columbian coca cartels.

When this happens, even the most obstinate puritans will have to face a hard fact of life that it's hopeless to outlaw victimless, pleasure-seeking behaviors, even when they are self-destructive. You *can* regulate. You can

tax, preach, and persuade. You can even heal. Recent, gradual success in the great-big-war-against-tobacco proves that patient temperance drives can do a lot of good, without trampling on the Constitution or a man's right to choose his own road to hell.

Nevertheless, there finally comes a point when a society would be well served to remember the admonition offered by Clint Eastwood at the end of *Magnum Force*: "A man's just got to know his limitations."

### The Wild Net Yonder

None of the technologies I've mentioned so far will change our lives anywhere near as much as the information revolution. It's a hot topic. Everyone is talking about huge data highways, linking home to office, to factory, to school, to the AI assistant you'll have strapped to your wrist. All the world's channels of communication, all the databases, all the libraries, will be merged into one universal network. Everything will be connected. Knowledge will be power.

Yet even on this new frontier we are already seeing the great skill some people apply to making nuisances of themselves. Today's Internet, for instance, is plagued by countless individuals who fly into abusive, self-righteous tirades called "flames" at the first sign of disagreement. Much like mental patients suffering from Tourette's syndrome, they blurt out angry vituperation, abandoning the editing process of common courtesy that took us thousands of years to acquire. (At last year's ALA conference, I gave these people a name—Net-Tourette's.)

Yet are these flammers all that different from the motorist who cut you off last week, flipping an obscene gesture and laughing at your frustration, safe behind his mask of anonymity? The problem may not be entirely due to the lack of visual cues common in face-to-face contact, as so many pundits have

insisted. Rather, flaming may be just one more symptom of a much bigger problem, which we'll take up later.

The information revolution is already affecting our political sensitivities. It used to be a truism that the candidate who was ahead in the polls would duck her opponent's calls for a debate. Recently, however, public attitudes of fair play seem to insist that debates be considered an essential part of the process, and being ahead is no longer an excuse for skipping out.

Television did this, and the process continues. Once upon a time, when a candidate published his or her income tax return, it could be dismissed as a stunt. But today that, too, is fast becoming the expected thing. (In France, a land of persnickety individualists, *all* tax returns are a matter of public record. They survive.)

Soon politicians will be required to keep all financial accounts open to public scrutiny on the Net. The new technologies of communication from talk shows to the Internet will enforce this new era of accountability, for better or for worse.

Nor are politicians alone under the hot spotlight glare. Think about how *police work* has been changed by the arrival of cheap video cameras. They have proven invaluable tools, assisting the forces of law in winning arrests and convictions. And they have also made the beat cop feel as if the whole world is watching over his shoulder.

Already local neighborhood-watch groups are being trained to maintain stakeouts, helping track the movements of gangs and drug dealers on local streets. Security cameras have reduced mayhem on school buses, and they help deter crimes on subway platforms.

It sounds benign, but where will it lead? What will we do when surveillance tech keeps getting smaller, niftier, and lower in price? When someone can pilot a remote-controlled drone the size of a housefly through a neighbor's window, shall we then

pass laws to keep such devices out of the local electronics store?

All of the anecdotes I've just mentioned have one thing in common. They point out the critical interaction we face, between *technology*, *accountability*, and *privacy*. I expect few topics to create more ferment during the coming decade. As Americans, we adore our privacy even more than freedom. Even more than we love our flashy techno-toys. So when we are told (as we will be, quite soon) that "privacy laws" will protect us against nosy intrusion by our neighbors, by businesses and government, we'll vote for them, standing up and shouting "Aye!"

But nothing will protect or save privacy. It's over.

The anonymity of urban life was one of the attractions that brought millions to vast cities, escaping the village busybodies back home. But that armor of seclusion amid multitudes is going to vanish in the near future, like the chimera it has always been, a passing illusion, a fluke in the history of human communal living. Soon we are going to return to the way of life of our ancestors, to the transparency of the village, whether we like it or not.

Let me ask you something. How many of you think that if we pass privacy laws it will actually prevent the rich and powerful from finding out anything they want to about you? What, no hands raised at all? I thought not.

Clearly, all "privacy laws" will accomplish is to provide common folk like you and me with a warm, fuzzy *mirage* of seclusion, while having the major effect of preventing *you* from finding out anything at all about the rich and the powerful.

What shall we do about those housefly-sized spy drones I spoke of earlier? Banish them from Radio Shack? Restrict their use only to the police?

Right.

Or, rather, over my dead body. If anyone is going to have the power to

spy on me, I want them to know that I just might have a housefly of my own watching them watch me!

That's how courtesy was enforced in the old village. People understood that it was in their own best interest to be polite. You quickly learned that the best defense against bad neighbors was to work at being a good neighbor yourself.

Soon that fellow who laughed as he rudely cut you off on the freeway won't be able to hide behind a shield of anonymity anymore. The kid who swipes an apple from a shouting fruit vendor can expect to get a phone call on his wrist phone before he runs more than a block away. Would-be burglars will have to be *awfully* clever when cheap video cameras in any home can be automatically linked to the police, in real time. And the inveterate flamer on a computer bulletin board will find his system plagued by "courtesy worms," a type of immune-response software, sicked onto him by members of an offended Net community.

In the village, it wasn't fear of retribution that kept you from behaving rudely, callously, toward your neighbors; it was the sure knowledge that someone would tell your *mother* and bring shame to your family. Tomorrow, when any citizen has access to the universal database to come, our "village" will include millions, and nobody's mom will be more than a fax call away.

I reiterate; it's over. It was fun while it lasted, living on these city streets amid countless, numberless fellow beings, not knowing any of them at all. It was also lonely. Today, you read about old folks found dead in their apartments, months, even years, after anyone had last seen them alive. That won't happen anymore when the village returns. Busybodies will gossip, but you'll be able to leave your doors unlocked. Everyone will know how much you paid for your nose job and what videos you rent, but you and your kids will have friends in

every part of the world, whom you met through shared interests on the Net. And when you travel, those friends will pick you up at the airport with wide-open arms, even though you never laid eyes on them before in your life.

Perhaps, after all is said and done, most of us will even decide that it's better that way. Better to know our neighbors (in their multitudes) than to live a fiction, a lie, of splendid, lonely isolation.

As if we're going to have the slightest choice in the matter.

## Freedom and Politics on the Net

Let's talk about some other possible features of that information network of the future.

Here is one of my favorites. If we plan in advance, the myriad, multitudinous data streams caroming back and forth across cyberspace may allow room for adding something called tag commentary onto most messages. I am talking about a few parasitic bits affixed to the end of nearly every data stream, within which each reader or recipient can append a brief comment, perhaps as succinct as a plus or minus sign, signifying "thumbs up" or "thumbs down." These blips would be mere ephemera, of no *official* value, useful only in sampling the reactions of Net users to a particular tract or posting.

Now I'm certain the people in charge of the information pipelines will hate this idea, because it's going to add to an already overwhelming problem of bit-rate loading, yet this tagging could be used for any number of amazing things. Users may, for instance, put down flammers and diatribe purveyors without ever having to engage the perpetrators in direct mudslinging. Of course, some Net-Tourrete's will take pride in achieving the highest possible negative scores, but most users will simply adjust their

access-sieves to filter out anything tagged below, say, minus nine thousand, and so most of us will escape the swill, except when we choose to go slumming.

On the positive side, tagging might help elevate some particularly insightful missive, one posted by some unknown person, just once, in a dim corner of the Net, so that it might percolate upward out of obscurity, to reprinting, reposting, and perusing by ever-greater numbers, simply by virtue of the accumulated number of enthusiastic "nods" it receives along the way. In this way, the moguls who own vast media empires won't have a mortal lock on what we see or think. Rather, good ideas and good art will have an alternative route to wide dispersion through a system of simple, anarchic, popular merit.

One result of tagging I would especially like to see is the potential for assessing pompous would-be pundits with *credibility ratings*. For instance, picture a typical twenty-first-century news reporter coming onscreen. Underneath his talking head, you see an accumulated *score* (or several scores, one compiled by, say, analysts at *Consumer Reports*, and another gathered from viewers, watching in real time, reacting with thumbs up or down). These little numbers would serve to show how trustworthy or believable viewers are finding the product (whether a sales pitch or news or commentary) that the reporter is trying to "sell."

Imagine these credibility ratings *changing* in real time. Envision the sweat popping out on our ace authority's brow, as his score rapidly plummets before his eyes!

Now picture a *politician* having such a figure flashing away at the bottom of her TelePrompTer!

Yes, the bad side of all this is obvious. It could serve to elevate the level of debate or to debase and homogenize it, depending on how the system is used. Which is why I'm raising such possibilities, both here and in my novels. It may be too late to *prevent* these

trends, but it is by no means too soon to begin thinking about possible consequences.

Indeed, this is surely not the first time something similar has happened. History warns us to be wary every time a new communications technology arrives, because while some try to use it to uplift humanity, others skillfully apply the new medium to the oldest of magical arts—manipulating others.

The old aristocrats in 1930s Germany, for instance, thought they could control Hitler, because they owned the newspapers. He went *around* the press, reaching vastly greater masses with the new hypnotizing power of radio and loudspeakers, which, to people newly exposed to them, seemed to amplify the user to greater than life . . . to near godhood. In a parallel vein, the Ayatollah Khomeini bypassed the shah's monopoly on radio and TV simply by smuggling into Iran one single cassette tape per week. Within days, that cassette had been duplicated a thousand times, to be circulated and played aloud in every mosque. Fax machines came close to serving the same insurrectionary function in China during the Tiananmen uprising, which nearly toppled the old Communist aristocracy.

All of these fruits of science have proven both empowering and potentially manipulative of the common man and woman. And the drama of our yin and yang, love and hate, relationship with technology continues.

What will be the consequences when, as some predict, the personal computer is so cheap that the average citizen of the Third World owns one and has greater access to *data* than clean water?

We are bound for interesting times.

## Slavery in the Future

With one noted exception, the social structure of nearly every human civi-

lization has been pyramidal in shape, with a few aristocrats lording it over great masses of peasants whose function was to serve their masters' whim. The greatest exception to this historical rule is our own. Modern Western society appears to be shaped like a *diamond*, with the vast majority living fairly decent lives in the middle. What aristocracy we have must, supposedly, *earn* its place only slightly above the rest of us, and earn it again with each generation. It may be impossible to eliminate poverty entirely, but our poor should have reasonable hope to join the center, if not personally, then through their children. The diamond has a name: the American dream. And even after the attempted aristocratic coup of the 1980s, this society still has the astonishing shape of a middle much larger than the bottom. A social structure in which the well-off still outnumber the poor. A shape that, despite all the work we have yet to accomplish, is unprecedented in humanity's long history of complex civilization.

Now before we get smug, let's reveal a secret truth. The diamond exists, but only for the aristocratic class of *human citizens*. Underneath that wedge of privilege, supporting it, crouches an invisible pyramid of *slaves*. Vast numbers of drone workers, without rights or entitlements, who toil at the beck and call of even our humblest proletarians. The slaves are called *machines*, billions, trillions of them, depending on how you define the word . . . or perhaps numberless beyond counting, if all software entities are included, and they are the modern helots on whose backs the human master race rides toward a splendid tomorrow.<sup>1</sup>

This idea is not new. Science fiction is rife with foretellings of possible consequences of this new social pyramid. In his novel *Mockingbird*, Walter Tevis depicted a decadent humanity, "served" into somnolent oblivion by a resentful machine class. Cordwainer Smith evoked sympathy in his works

for the slaves we seem bound to create, out of both machinery and altered animal life. The works of Jack Williamson send chills up one's spine, with visions of how robots might reinterpret their fundamental commandments to "protect" men and women once their artificial intelligence surpasses our own.

On the opposite extreme, we have vistas of bliss and fulfillment in Vernor Vinge's novels, which feature something he calls the "Singularity," a near-future event in which human knowledge, and our ability to "tie into" the culture-wide database, start to expand geometrically . . . asymptotically . . . with our collective skill rocketing higher and higher, faster and faster, as we synergize with our synthetic agents, so that, within mere months, every sentient being on earth attains godhood and leaves.

This is more than just philosophical mauling about history and destiny. The metaphor of servant and master is vitally relevant to the way millions today use their interface with the growing, worldwide Info Net. For like television, and radio before it, the new technology has the capability of metastasis, of turning round and devouring the body it was meant to serve and taking over the life of the "master" it was designed to benefit.

Be honest. How many of you in this audience have grown addicted to your computer bulletin boards, your interest groups and data access nodes? How many hours a day do you devote to this new artificial world? Hours taken away from your real work, your real world relationships? True, the opportunities offered by the Net are glorious. Every day we see the upside described in glorious detail in magazine articles, on TV, and, especially, on the Net itself. But surely you'll admit that it can also become a tar baby, a parasite, sucking hundreds, even thousands, of hours out of your valuable life span—in effect, hours sucked away by a machine.

As for me, I'm a curmudgeon. De-

spite the optimistic depictions I present in *Earth* and other novels of a future Info Net that ennobles humanity, I remain deeply suspicious of its present incarnation. I own the machine; the machine does not own me! So I use e-mail sparingly and stay away from Net groups that might draw me in with tantalizing but time-devouring discussions.

Am I denying myself wonders? Probably. But not for long. To properly utilize the miracle of worldwide data access, I am waiting for a few new slaves.

Prototypes are already out there, called software "agents." (In *Earth* I call them "ferrets," "hounds," and even "hey you.")

Picture the work space of tomorrow. You *appear* to be all alone at your desk. But entities await your bidding, patiently, ready to act on any whim.

You'll say, "Ferret! I want some information!" and your home computer will quickly reply, "Yes, boss. What data do you want me to fetch for you?"

"Go get me five hundred words, written in the style of Connie Willis, about Scandinavian sauna architecture and associated mating rituals, okay? Make it funny, and accompany with illustrations in the style of Vargas. Chop, chop."

"Yes, boss," the voice replies. And off a little messenger program scurries outward, into the vast sea of the Net, to fish for what you need and prepare it according to your tastes. Now *that's* when this particular skeptic will become a true convert. It's when Vernor Vinge's vaunted Singularity might be on the verge of taking off, for our capabilities will then start to multiply as never before.

Unfortunately, there will also be a dark side to this trend toward loyal servant entities in the form of privacy screen guardians and personalized data sieves—the insularity of perfectly isolated worlds of the mind. Some people will inevitably program their home units to admit only those

TV programs, only those shows and magazines, that agree with opinions they already have. We've already seen this trend in channels devoted to certain ethnic and religious groups and in the cult followings of talk-show hosts, such as Rush Limbaugh.<sup>2</sup>

Will "connecting" technologies only serve to divide us in the end, fragmenting us into a myriad acrimoniously bickering tribes? Could it be that those working on the Tower of Babel were thus cast into confusion, cursed to speak countless indecipherable tongues, by means of something as useful and promising as the Net?

There are some ways in which our tendency to fractionate may prove ennobling. As we form interest groups and join bulletin boards covering each subject under the sun, gathering together from all corners of the globe every hobbyist who is interested in, say, miniature toy soldiers, or Chinese porcelains, or origami, it is fast becoming clear that no single thing of value from the past or present will ever again be completely lost. If it has any intrinsic merit at all, it will attract some band of eccentric aficionados who will turn it into an avocation, a pastime with an ardent cult following that crosses all borders and time zones. Already, the trend can be seen flourishing. There are more spinning wheels and floor looms being used today by hobbyists than existed in the nineteenth century. Horses and blacksmiths abound in the era of the gasoline engine. There are more *sword makers* in America today than lived in the entire Middle Ages, making blades that could have sliced through Charlemagne's best armor like butter. Moviemakers can once again film vast war epics in the United States, filling battle scenes with unpaid extras who provide their own perfect Civil War uniforms, marching and "fighting" with a realism that could only arise out of the most passionate of hobbies.

And on the Net there thrive special corners where people seek to share

knowledge and grow in craftsmanship through ten million pastimes.

So which destiny will be ours? At one extreme, the future depicted by the cyberpunks, while containing no end of dazzling wonders, will be dominated by criminals, aristocrats, and vast, faceless bureaucracies. If the dourest cyberpunks are right, then our institutions are irredeemably corrupt and our fellow citizens irremediably stupid and controllable, as depicted in Oliver Stone's ugly television drama "Wild Palms." If so, if our fellow citizens are as witless as Stone and his ilk clearly believe, then we certainly shall deserve nothing better than the downer tomorrows those authors depict.

On the other hand, some of the incisive works of Orwell, Huxley, and later authors such as Brunner, Sterling, and Sheffield may affect our thinking in more useful ways, inspiring wary caution while helping us to avoid the very mistakes depicted in their novels and films. The *self-preventing prophecy* is a venerable subgenre of SF.<sup>3</sup> A dire warning can be valuable indeed, so long as it offers implicitly or explicitly the possibility of a way out.

## Where It All May Lead

As I've said before, prediction is hazardous and often misleading. While some distant vistas may terrify us into avoiding error, others might serve as beacons to guide us forward, enkindling hope to carry us across the rough days ahead.

Last year, Hans Moravec dazzled the crowd at the last ALA conference with his notions of how, if we do succeed in building a decent civilization, our descendants may be indistinguishable from gods in their ability to contemplate vast numbers of possible realities or to create simulated cororealities down to the minutest detail. If this does come about, what

will our descendants do with such power?

One of their hobbies, Hans predicted, will be to concoct accurate models or re-creations of past events. Models so perfect that each of the simulated players in a scene will actually think synthetic thoughts, with all the complexity and subtlety one could desire, just as if the event were "real."

In fact, at this point I might ask, How do any of you out there in the audience know for certain that you and this room are *not* simulations? Perhaps we are all magnificently delineated personas, playing roles in a deluxe reenactment of one special, historic day, way back in the late twentieth century, when a couple of otherwise obscure sci-fi writers—Connie Willis and David Brin—uttered a few remarks that stimulated an idea in the head of *that* young man in the third row . . . an idea that later changed the world?

Okay, everybody, now I want you all to turn toward that wall over there and wave at our descendants! That's right. Maybe *that's* what's special about this moment, the aspect that gets this simulation replayed over and over in the sixtieth century as a minor classic. The first simulation that showed evidence of self-awareness and even tried to converse with its simulators!

What a classic.

We should be so lucky.

In conclusion, I want to return for just a moment to the issue that may have struck the most nerves when I raised it a little while ago. That is the issue of privacy and my suggestion that the era of leading anonymous lives, each of us dwelling almost alone amid the urban multitudes, is about to end. Laws intended to seal off our private information from prying eyes will only guarantee that those prying eyes must be rich and powerful in order to roam at will across our personal affairs. Those laws will also make certain you and I have no access

at all to information about the rich and powerful, except what they choose to share.

We can have either freedom *or* privacy, I predict. Alas, not both. Given that dichotomy, if the alternative is to give up liberty, my own choice is clear. As long as the playing field is even, and *everyone* is subject to the same rules, perhaps privacy should be exposed as the myth that it has always been.

Now I know this puts me in a tiny minority. And an unpopular one, I might add. Not only are civil libertarians on the march, demanding legislation to protect a phantom promise of data seclusion, but they are being aided by a strange alliance of conservative and business leaders. And a lot of hackers, too. Most of the cyberpunks, it seems, have decided to come down with full, self-righteous force on one side of this issue. Alas, the wrong side.

Let me quote briefly from the *Cyberpunk Manifesto*, by Eric Hughes:

Privacy is necessary for an open society in the electronic age. Privacy is not secrecy. A private matter is something that one doesn't want the whole world to know, but a secret matter is something one doesn't want anybody to know.

[Huh?]

Privacy is the power to selectively reveal one's self to the world. We must defend our own privacy if we expect to have any. We must come together and create systems which allow anonymous transactions to take place.

[Like those anonymous letter bombs that have recently been blowing up on innocent people's desks, I suppose. And computer viruses? But let's go on.]

People have been defending their own privacy for centuries with whispers, darkness, envelopes, closed doors, secret handshakes, and couriers. [Oh, good stuff. Admirable!] The technologies of the past did not allow for a strong privacy, but electronic technologies do. Let's extend all those other things, the whispers, the dark-



ness, the envelopes. We the Cyberpunks are dedicated to building anonymous systems. We are defending our privacy with cryptography, with anonymous mail forwarding systems, with digital signatures, and electronic money.

Over my dead body. And yours as well, I presume, since librarians like you in this audience are even more dedicated than I am to the open transmission and sharing of information.

Information. Fools predict that it will be the "money of the future"!

But it is not money.

It's not wealth.

It is *life*, and its unfettered course is as necessary to a healthy, free society as the open flow of air.

As far as I'm concerned (and this may sound odd coming from a fellow who grew up in Los Angeles), the *air* is too precious to become a commodity. Air should always be free.

And so should be the light.

*Note: If you enjoyed this article, you might be interested in attending the Imagineering Interest Group program in Miami, where two stellar science fiction writers, Gene Wolfe and Greg Bear, will speak.*

## Notes

1. Consider a simple, mundane case. You can all get a snack, in the middle of the night, faster than a pharaoh could by snapping his fingers and commanding a hundred servants. All *you* need is those two loyal, efficient slaves—your fridge and your microwave oven.

2. In fact, the greatest science fiction author ever to have lived may be Karl Marx. In the East he was taken as a religious leader, providing a dogmatic doctrine to serve as a semireligious focus for what will eventually come to be seen as just another massive peasant uprising—more virulent than most, but akin to countless others throughout history—which was betrayed (as usual) by the generation of revolutionaries themselves, who schemed to make themselves into aristocrats in the new order.

Where Karl Marx had truly historic effect was in the *West*, where his scenarios were read as terrifyingly believable *stories* about what might happen in the near future if present trends continued. In science fiction, we call this the "if this goes on" tale. And when such a tale is particularly terrifying, it galvanizes reform-minded individuals to act, to push for changes in the system. In this case, the changes wrought by liberal reformers were so drastic and far-reaching that the terrifying scenario depicted so believably by Marx—of an ever-narrowing social pyramid leading to

upheaval—no longer seems at all likely to come about. The ultimate irony is that Karl Marx may have been the chief destroyer of his own dream! If you credit that Franklin Delano Roosevelt and others yanked capitalism off its former track, onto one in which workers own a piece of the action and have a stake in capitalism's success, then it isn't all that far-fetched to fantasize that Karl Marx helped prevent the very social breakdown he predicted so vividly in *Das Kapital*.

Very few authors have ever probably predicted or *made* future events happen simply by writing words. But a great many may have played roles in *preventing* possible tomorrows. It is an intriguing suggestion. Of course, by its very nature, it is also a contention that cannot be proved. Still, I stand by my suggestion that Marx be reassigned to the category of a major sci-fi writer, despite the fact that millions took his works of fiction much too seriously. (Let that be a warning to the rest of us.)

3. Actually, people can be awfully forgiving of would-be prophets. One success counts more toward your reputation than a hundred failures. People still come up to me, claiming awe that I predicted the fall of the Berlin Wall back in 1987. No one bothers to chide me for also prophesying a solar-sail mission to Mars, a Nobel Peace Prize for Jimmy Carter, and at least *some* success for supply-side economics. Fortunately, those embarrassments have been conveniently buried and forgotten.

## The Wilson Abstracts Advantage

Wilson Abstracts give you the advantage of meeting all of your patrons' periodical research needs at a fraction of the price of full-text resources. The six new databases—covering the areas of science, technology, art, education, the humanities, and social sciences—are modeled after Wilson's popular *Readers' Guide Abstracts* and *Wilson Business Abstracts*, and enhance access to information for *all* researchers.

## Wilson Abstracts Are BETTER Abstracts

Written by professional abstractors, Wilson Abstracts set the standard. Each abstract ranges from 50 to 150 words in length and often contains all the needed information—including facts and figures—making it unnecessary to locate the original article. The detailed abstracts help preserve your periodical collection and speed research. Even when a given article is not available in your library, you can count on Wilson Abstracts to accurately convey the content, tone, and scope of the original text—every time!

## Wilson Abstracts Feature:

- ◆ Comprehensive indexing and abstracting
- ◆ Full, informative abstracts that speed research and often substitute for the original article
- ◆ Straightforward subject terms for one-stop searching
- ◆ Approximately ten years of indexing coverage in each database

## WILSONDISC Features:

- ◆ Monthly disc updates for most databases
- ◆ Unlimited online access to the corresponding WILSONLINE database
- ◆ Generous networking policies for single and multiple institutions
- ◆ The privilege of keeping all discs
- ◆ Three WILSONDISC search modes that accommodate all users.

## Get the Wilson Advantage

Call **800-367-6770** today for your 90-day Wilson Abstracts trial on WILSONDISC. Wilson Abstracts are also available online and on tape.

### **Wilson Applied Science & Technology Abstracts**

*Abstracting and indexing of 400 key science, trade, and technology periodicals. Coverage includes everything from aeronautics and space science to transportation. Monthly disc updates. Updated twice weekly online. Available May 1994.*

### **Wilson Art Abstracts**

*Abstracting and indexing of 268 leading domestic and foreign periodicals, yearbooks, and museum bulletins. Quarterly disc updates. Updated twice weekly online. Available July 1995.*

### **Wilson Education Abstracts**

*Abstracting and indexing of 400 core international English-language periodicals, yearbooks, and monographic series covering all levels of education, from preschool to post-graduate. Available July 1995.*

### **Wilson General Science Abstracts**

*Abstracting and indexing of 140 core science journals and the science section of The New York Times. Monthly disc updates. Updated twice weekly online. Available May 1994.*

### **Wilson Humanities Abstracts**

*Abstracting and indexing of 350 core periodicals on literature, religion, philosophy, history, and related topics. Quarterly disc updates. Updated twice weekly online. Available March 1995.*

### **Wilson Social Sciences Abstracts**

*Abstracting and indexing of more than 400 English-language periodicals in all areas of the social sciences, from anthropology to urban studies. Monthly disc updates. Updated twice weekly online. Available March 1995.*

### **Also Available...**

*Readers' Guide Abstracts  
Wilson Business Abstracts*

# WILSON ABSTRACTS

 **THE H.W.  
WILSON  
COMPANY**  
Sharing Knowledge  
Since 1898

**Branching Out to Meet Your Growing  
Research Needs with Six NEW  
Wilson Abstract Databases**

## Mapping 1990 Census Data

Patrick Yott

*Depository libraries can now receive electronic data products from the Bureau of the Census. Included in this category are the Summary Tape Files from the 1990 Census, the 1987 Economic Census, County Business Patterns, as well as TIGER, the Census Bureau's geographic map database. Armed with these products, librarians can offer their patrons a wide variety of choices for viewing data. Most exciting is the possibility of presenting data graphically. Using a variety of software packages, the University of New Hampshire library has begun to offer its patrons a variety of data products, ranging from printed census volumes to customized spreadsheets, databases, and maps. The process employed in producing a thematic map using the 1990 Census is described.*

The 1990 Census was special: not only was 1990 the bicentennial of the first U.S. Census, but more important (at least to librarians), the 1990 Census was the first census to have electronic data distributed to the general public via the Depository Library Program. For the first time, individuals and small businesses had access to data that previously had been unaffordable or unusable. For some librarians, however, jubilation soon turned into sobering panic as they realized that they would be called upon to function as intermediaries between the Census Bureau compact discs and the end user. This was new territory, and it looked dangerous.

The University of New Hampshire (UNH) library has been fortunate enough to receive several grants to assist in developing a "government data center." Two grants were awarded by an Undesignated Gifts Fund at the university and provided

for hardware and software. A third grant was awarded by Strategic Mapping, Inc., in the form of a copy of the Geographic Information System (GIS) software package, ATLAS GIS.

To illustrate how one can extract data, manipulate it, convert TIGER files into something usable, and finally merge the data with a map file (see figure 1), the following hypothetical query has been devised. The patron is working on an assignment for a graduate course in marketing and needs to analyze the black population of Essex County, Massachusetts, to target advertising. The variables she has chosen to work with are:

- a. Black population
- b. Black population as a percentage of total
- c. Percentage of blacks completing college
- d. Per capita income for blacks

The student needs maps generated for variables *b*, *c*, and *d*. This tutorial will demonstrate a method to do the following:

- a. Prepare a base map from TIGER files.
- b. Use EXTRACT to create a series of outfiles.
- c. Combine outfiles using dBASE.
- d. Convert a dBASE file into a Lotus 1-2-3 file to perform math functions.
- e. Convert the Lotus file back into dBASE.
- f. Produce a thematic map of this data using ATLAS GIS.

### The Census Files

TIGER stands for Topologically Integrated Geographic Encoding and Referencing system. These files provide a computer-readable map/geographic database for the entire United States. When used with a GIS package, TIGER files form the basis for an endless

variety of maps on which to plot census and other data.

STF (summary tape file) data are machine-readable data sets that provide the basis for the printed census volumes. Two major STF data sets are being distributed for the 1990 Census, STF 1A and STF 3A. STF 1A data sets contain aggregate data obtained from the 100 percent count (short form) of the 1990 Census. The STF 3A data sets contain aggregate data from the sample count (the long form that was answered by one in six households). Data found on STF 1A files are core demographics, while STF 3A details education, ancestry and migration, occupational status, income and poverty, as well as a wide variety of housing characteristics.

### Software

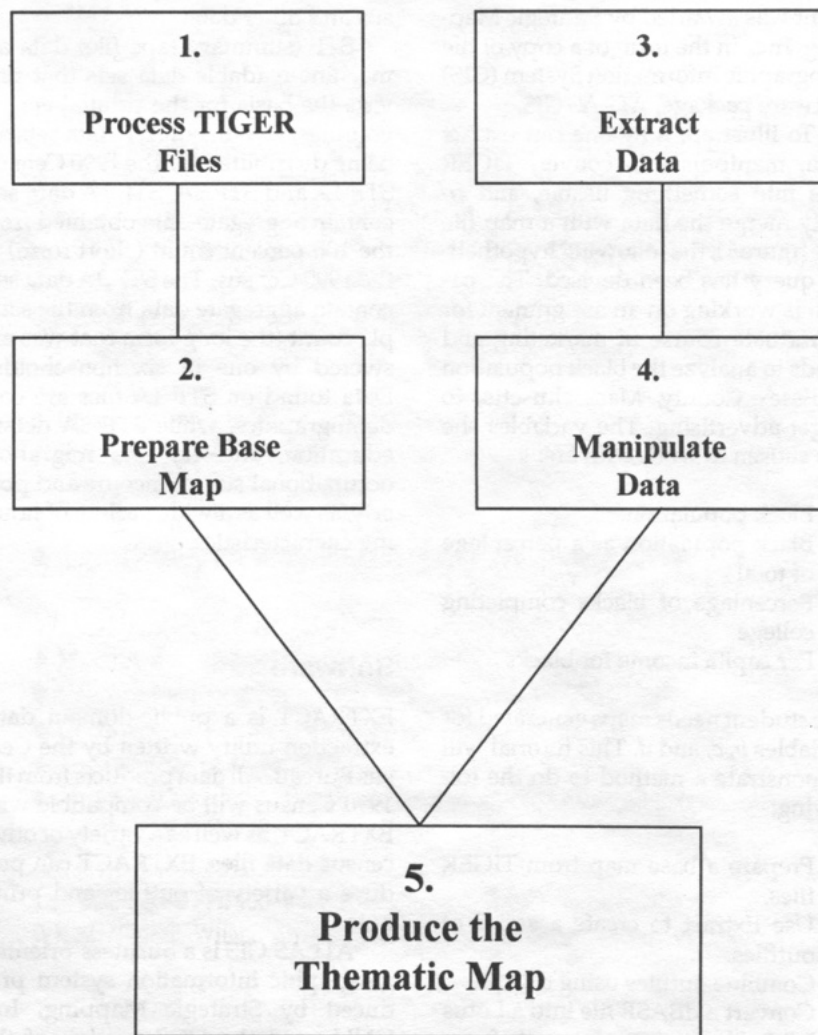
EXTRACT is a public-domain data-extraction utility written by the Census Bureau. All data products from the 1990 Census will be compatible with EXTRACT, as well as a variety of other census data files. EXTRACT can produce a variety of outfiles and printouts.

ATLAS GIS is a business-oriented geographic information system produced by Strategic Mapping, Inc. UNH used the DOS version of the program, but a Macintosh version is also available. To use TIGER files with ATLAS GIS, a secondary, polygonizing program was necessary, and the library chose TGRBdry Translator (a product of Micro Map & Cad).

The UNH library uses both dBASE III Plus and Lotus 1-2-3 to perform a variety of administrative tasks. When the Census Bureau began distributing most of its data in dBASE format, UNH was already versed in using the

---

Patrick Yott is Documents Librarian, Dimond Library, University of New Hampshire, Durham.



**Figure 1**  
The Mapping Process

software. ATLAS GIS requires that attribute (data) files be in dBASE format (\*.dbf). Lotus 1-2-3 was the choice for performing mathematical operations, although other spreadsheet programs could have been used. The ease of migrating between the spreadsheet and dBASE should be considered when choosing a spreadsheet package. (Lotus 1-2-3 and dBASE have built-in commands to facilitate data exchange.)

### Preparing the Base Map

TIGER files are prepared by county, with twelve file types produced for each county. Files are named by the FIPS (Federal Information Processing Standards) codes for the state and county they represent (in the present example—state 25 and county 009). File extensions identify the type of TIGER file. To create a base map for the project, only three basic files were needed:

tgr25009.f41—The basic data records file  
 tgr25009.f42—The shape coordinate file  
 tgr25009.f43—The additional polygon file

TGRBdry can be used in an interactive mode where it prompts the user for information. Operating the program in command mode as part of a batch file, however, allows unattended processing, and this is helpful because processing can be lengthy depending on the amount of data.

A simple batch file (essex.bat) can be prepared as follows:

```

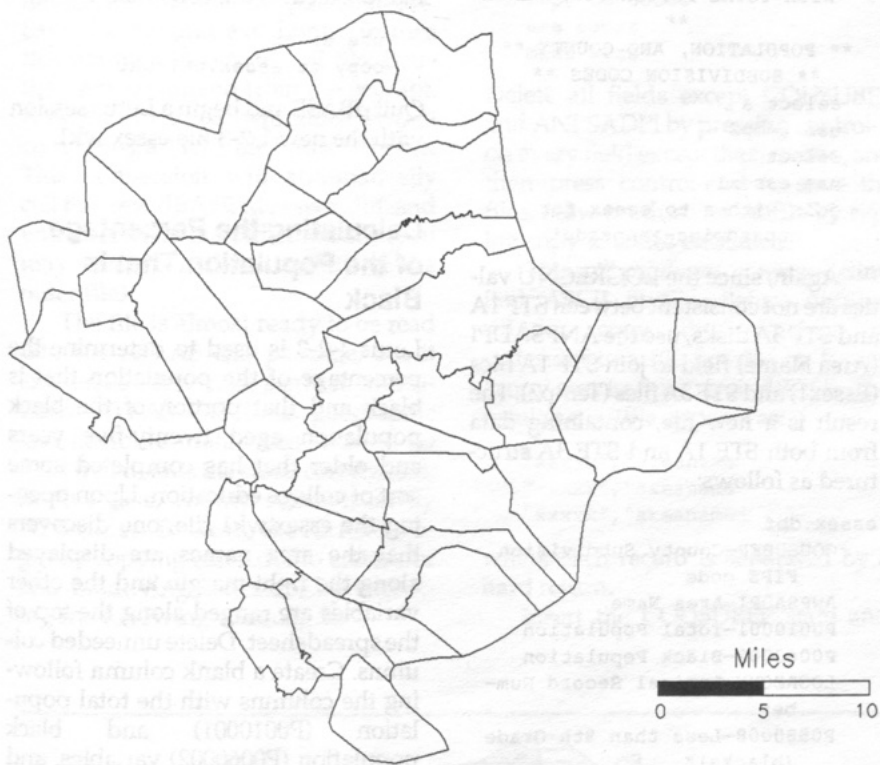
copy 1:\25\009\*.f41
copy 1:\25\009\*.f42
copy 1:\25\009\*.f43
tgrbdry tgr25009.f41
      tgr25009.f42 tgr25009.f43 1
      8 on bna essex ie essex.bna
      essex.agf /names 3
  
```

The first three lines copy the needed TIGER files from the CD-ROM to a hard drive; operating from a hard drive greatly speeds up the program. The fourth line activates the TGRBdry Translator program. Parameters for TGRBdry Translator are:

```

tgrbdry—starts the TGRBdry Translator
tgr25009.f41—identifies the basic data records file
tgr25009.f42—identifies the shape coordinate file
tgr25009.f43—identifies the additional polygon file
1—identifies the projection (1=lat/long.)
8—identifies the boundary choice (MCD\CCD FIPS)
on—turns the lake switch on
bna—instructs the program to create an ATLAS usable output
essex—filename for the output file
  
```

The final line of the batch file activates a second translation program (ATLAS IMPORT/EXPORT), which converts the essex.bna file into an ATLAS GIS geographic file (essex.agf) that produces the base map pictured in figure 2.



**Figure 2**  
Base Map of Essex County

## Extracting Data

This project will require data from both STF 1A and STF 3A. STF 1A will provide data for total population and black population, and STF 3A will provide the educational attainment and income data. It is important to note and remind the patron that these two data sets are based on different samples; the 1A data being the 100 percent (short form) count and the 3A data being based on the one-in-six (long form) sample counts.

While the Census Bureau provides GO software with most compact discs, it is too inefficient for all but the simplest queries, and UNH relies, instead, on Extract and dBASE to work with the CDs. It is helpful in understanding Extract to conceptualize the

program's actions in terms of *x* and *y* coordinates. Option 1 (Select ITEMS) places variables along the *x* axis. Option 2 (Select RECORDS) determines the cases that will be displayed along the *y* axis. In the present example, population, educational attainment, and income form the *x* axis, and the county subdivisions (the census term for "towns") in Essex County should be displayed down the *y* axis.

Extract creates files as listed below.

### From Summary Tape File 1A:

essex1.dbf containing the variables:  
 COUSUBFP—County Subdivision FIPS Code  
 ANPSADPI—Area Name  
 P0010001—Total Population  
 P0060002—Black Population

### From Summary Tape File 3A:

essex2.dbf containing the variables:

LOGRECNU—Logical Record Number  
 ANPSADPI—Area Name

essex3.dbf containing the variables:

LOGRECNU—Logical Record Number  
 P0580008—Less than 9th Grade (Black)  
 P0580009—9th to 12th Grade/No Diploma (Black)  
 P0580010—High School Graduate (Black)  
 P0580011—Some College, No Degree (Black)  
 P0580012—Associate Degree (Black)  
 P0580013—Bachelor's Degree (Black)  
 P0580014—Graduate or Professional Degree (Black)

essex4.dbf containing the variables:

LOGRECNU—Logical Record Number  
 P115A002—Per Capita Income (Black)

Create these four database files as follows. Under OPTION 1 place an *x* next to each desired field; this will represent the values on the *x* axis. Under OPTION 2, select records by placing an *s* beside SUMLEV (summary level) and CNTY (county). The program then asks for the summary level—select level 060—STATE—COUNTY—COUNTY SUBDIVISION. It then will prompt for the appropriate county. Choose Essex. All of these choices will appear on a selection menu.

The data are located on separate CDs and in separate files. It will be necessary, then, to combine the various output files to create a single file that contains all the data (provided the data do not exceed the maximum number of fields allowed by dBASE—128 for dBASE III Plus; 256 for dBASE IV). Normally the key to joining the outfiles is the LOGRECNU field,

which is present in every field on the census CDs. Unfortunately, the LOGRECNU fields do not match between the STF 1A and STF 3A files, so it is not possible to join the outfiles from the two sources by this key. For this reason, LOGRECNU was not even included in the one outfile from STF 1A. Matching between STF 1A and STF 3A was done on the common field of area name (ANPSADPI).

Joining files in dBASE requires the repetition of a simple, five-line program:

```
select a
use <filename>
select b
use <filename>
join with a to <newfile> for
<field>=a-><field>
```

The select statements create work spaces so that you can open up more than one file at a time. The JOIN command looks for the appropriate field (usually LOGRECNU) and connects the file in the *a* space with the file in the *b* space only if the values in the appropriate field match. It is important that there is a one-to-one correspondence for values in the field linking the two database files or the program will give unwanted results. If done correctly, the new file should have no more records than the smaller of the two original files.

In the present example the program would read as follows:

```
** COMBINE PER CAPITA INCOME **
** (BLACKS) WITH EDUCATIONAL **
** ATTAINMENT (BLACKS) **
select a
use essex4
select b
use essex3
join with a to temp for
logrecnu=a->logrecnu
** COMBINE THE PREVIOUS SET **
** WITH AREA NAME **
select a
use temp
select b
use essex2
join with a to temp2 for
logrecnu=a->logrecnu
```

```
** COMBINE THE PREVIOUS SET **
** WITH TOTAL POPULATION, BLACK **
**
** POPULATION, AND COUNTY **
** SUBDIVISION CODES **
select a
use temp2
select b
use essex1
join with a to essex for
anpsadpi=a->anpsadpi
```

Again, since the LOGRECNU values are not consistent between STF 1A and STF 3A disks, use the ANPSADPI (Area Name) field to join STF 1A files (Essex1) and STF 3A files (Temp2). The result is a new file, containing data from both STF 1A and STF 3A structured as follows:

```
essex.dbf
COUSUBFP—County Subdivision
FIPS code
ANPSADPI—Area Name
P0010001—Total Population
P0060002—Black Population
LOGRECNU—Logical Record Number
P0580008—Less than 9th Grade
(blacks)*
P0580009—9th to 12th
Grade/No Diploma (blacks)*
P0580010—High School Graduate
(blacks)*
P0580011—Some College, No
Degree (blacks)*
P0580012—Associate Degree
(blacks)*
P0580013—Bachelor's Degree
(blacks)*
P0580014—Graduate or Profes-
sional Degree (blacks)*
P115A002—Per Capita Income
(blacks)
```

At this point, all the data have been extracted from the CDs in a single dBASE file. The format of the data, however, does not match what the patron really needs, as she wanted some data presented as a percentage. It was decided to convert the dBASE file into Lotus 1-2-3 to perform the mathemat-

\*Only persons twenty-five years and older are counted in variables concerning educational attainment.

cal operations. The following command executes the conversion:

```
use essex
copy to essex.wk1 wks
```

Quit dBASE and begin a Lotus session with the new 1-2-3 file essex.wk1.

## Calculating the Percentage of the Population That Is Black

Lotus 1-2-3 is used to determine the percentage of the population that is black and that portion of the black population aged twenty-five years and older that has completed some sort of college education. Upon opening the essex.wk1 file, one discovers that the area names are displayed along the right margin and the other variables are ranged along the top of the spreadsheet. Delete unneeded columns. Create a blank column following the columns with the total population (P0010001) and black population (P0060002) variables, and label the column "Percent." Enter a command to divide the figures in the "Total Population" column by those in the "Black Population" column and multiply them by 100, yielding a column of percentages of blacks in the population for the area names (ANPSADPI) ranged along the right margin.

## Calculating the Percentage of Blacks Twenty-Five and Over with College Degrees

Insert a blank column before the column containing the variable associate degree (P0580012) and label it "College." Enter a formula to divide the sum of blacks over twenty-five years of age with an associate degree or higher (variables P0580012, P0580013, and P058014) by the sum of the total population over twenty-five (variable P0010001) and multiply the quotients by 100 to arrive at percentages.

The essex.wk1 spreadsheet now has all the data that the patron needs. Save the file and exit Lotus. To turn this file back into a dBASE file, issue the trans command from the prompt. A series of choices will guide you in converting a \*.wk1 file into a \*.dbf file. This conversion will automatically call the new dBASE file essex.dbf, and will overwrite the old essex.dbf (you may wish to make a backup of the older file).

The file is almost ready to be read into the ATLAS GIS package, but it still needs a few more slight modifications. ATLAS GIS links geographic and attribute (data) files with an ID field that details the state, county, and, in this case, county subdivision by FIPS codes. To create the ID field, begin by copying essex.dbf to codes.dbf and modifying codes.dbf by executing the following commands:

```
use essex
copy to codes
use codes
modi stru
```

Delete all fields except COUSUBFP and ANPSADPI by pressing control-u on every field except the first two, and then press control-end to save the files. Now produce an outfile by typing copy to codes delimited.

This will produce a comma-delimited ASCII outfile. Before quitting dBASE, copy the structure of codes.dbf to a file fips.dbf by typing "copy stru to fips." Quit dBASE. The resulting outfile appears as:

```
"xxxxx", "areaname"
"xxxxx", "areaname"
"xxxxx", "areaname"
```

where each record is separated by a hard return.

Insert the COUSUBFP state and

county codes 25009 after the initial quotation mark on each line, so that the file appears as follows:

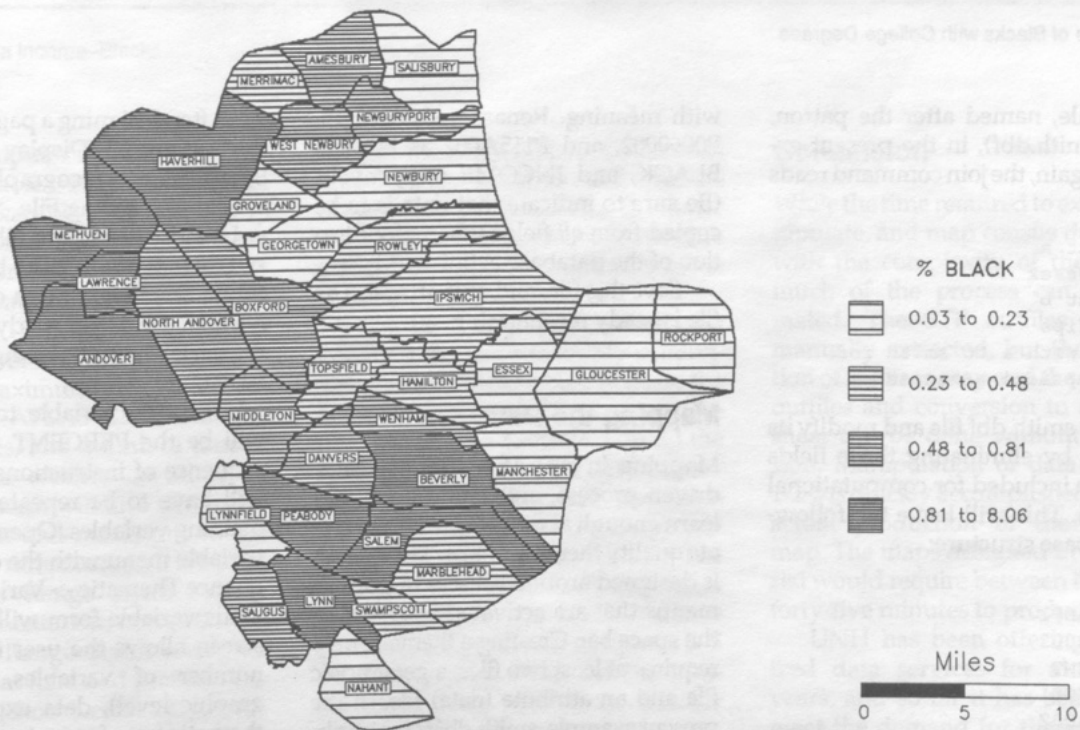
```
"25009xxxxx", "areaname"
"25009xxxxx", "areaname"
"25009xxxxx", "areaname"
```

Save the edited file as an ASCII file called fips.txt. This file is ready to be imported into the fips.dbf file.

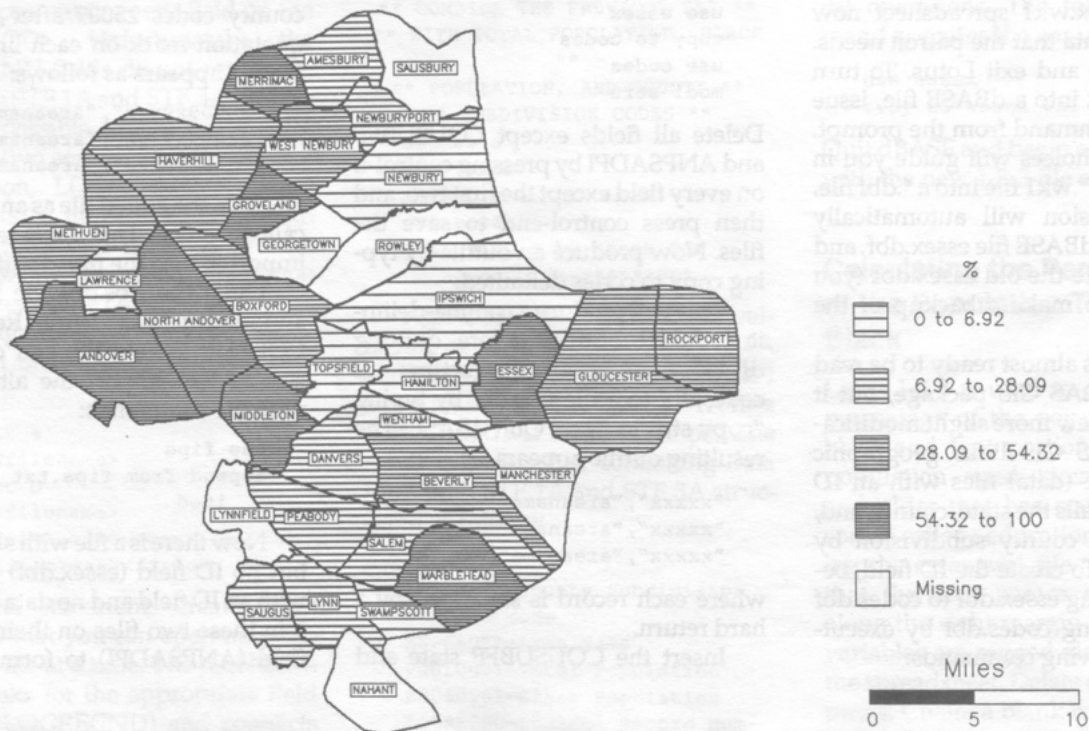
Open the fips.dbf file in dBASE and modify its structure. Rename the COUSUBFP field ID, and change its size to ten. Import the altered data with the commands:

```
use fips
append from fips.txt delimited
```

Now there is a file with all the data, but no ID field (essex.dbf) and a file with an ID field and no data (fips.dbf). Join these two files on their common field (ANPSADPI) to form the final



**Figure 3**  
Black Population of Essex County



**Figure 4**  
Percentage of Blacks with College Degrees

dBASE file, named after the patron, Smith (smith.dbf), in the present example. Again, the join command reads as:

```
select a
use essex
select b
use fips
join with a to smith for
  anpsadpi=a->anpsadpi
```

Open the smith.dbf file and modify its structure by eliminating those fields that were included for computational purposes. This will leave the following database structure:

```
ID
ANPSADPI
P0060002
PERCENT
COLLEGE
P115A002
```

It is helpful to rename the remaining coded fields to give them names

with meaning. Rename ANPSADPI, P0060002, and P115A002 as NAME, BLACK, and INCOME respectively. (Be sure to indicate that data is to be copied from *all* fields when modification of the database is finished.)

Save this file, and quit dBASE. The file is ready for mapping.

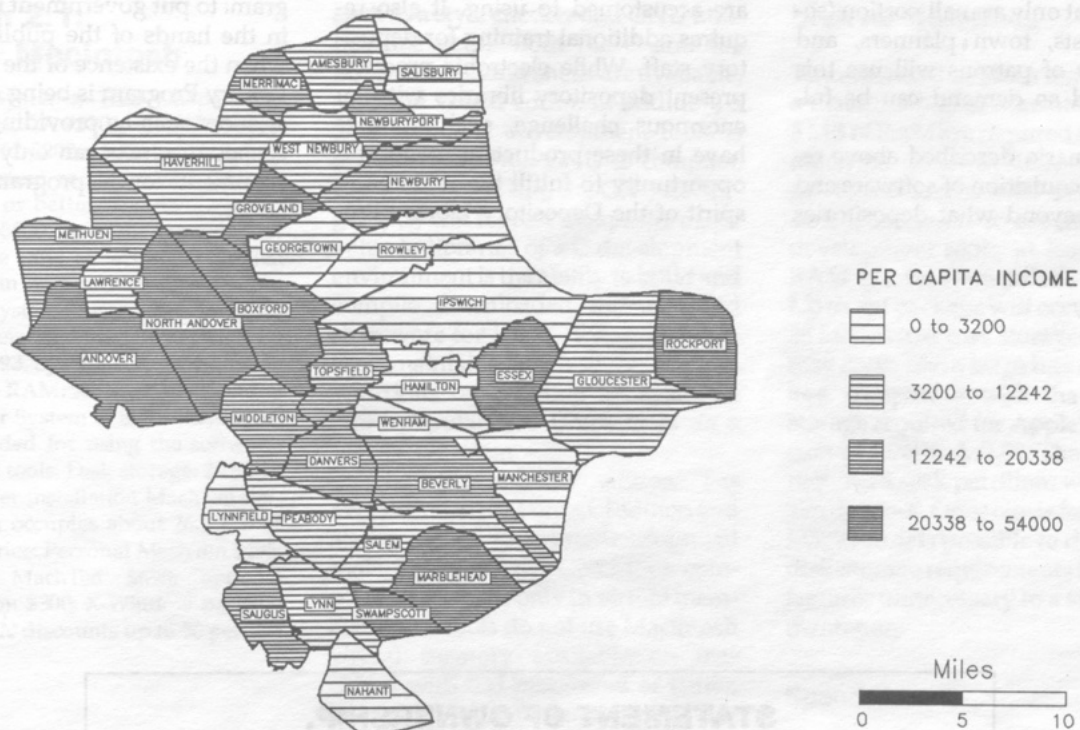
### Mapping the Data

Mapping in ATLAS GIS is a menu-driven process, and one can quickly learn enough about the process to create quality thematic maps. The system is designed around a series of pop-up menus that are activated by pressing the space bar. Creating a thematic map requires at least two files, a geographic file and an attribute (data) file. In the present example, smith.dbf is the attribute file and essex.agf (the base map created earlier) is the geographic file.

After designing a page layout with the command Display -> Page -> Setup, open the geographic and attribute files by typing File -> Geographic -> Use (use essex) and File -> Attribute -> Use (use smith). The base map (see figure 2) will be drawn on the screen. Now all the data is ready for choosing the variables and placing them on the base map.

The first variable to be mapped will be the PERCENT variable. The sequence of instructions follows and will have to be repeated for the remaining variables. Open the thematic variable menu with the command sequence Thematic -> Variable. The thematic variable form will appear. This screen allows the user to choose the number of variables, layer (geographic level), data expression, and thematic type for up to two variables. One variable will be mapped at the default geographic level of the base





**Figure 5**  
1989 Per Capita Income—Blacks

map (the computer will call it region). The first variable is PERCENT, and for the present example a ranged fill map is chosen. Once these choices have been made, press F10.

The thematic-setting form should now appear. The user can determine the number of quantiles, the minimum and maximum values, and the range styles. ATLAS GIS defaults to four quantiles, and a limit of four or five quantiles ensures an attractive map. The system will also determine the minimum and maximum values as well as the range values; it is best to accept them. Range styles require some experimentation, however. Pop up the Edit Range Styles menu by moving to that field and pressing the space bar to choose color or pattern scheme.

When the design of the map is completed, press F10 again to re-

turn to the main ATLAS GIS menu, and the base map should still be displayed. To draw the thematic map, press View -> Redraw. The screen should refresh, and a four-color map should appear.

The base map does not contain labels for the various county subdivisions. To add these labels, activate the layer settings menu by pressing Display -> Layer -> Settings. Pop up the Label Settings menu. For label expression choose ANPSADPI and then choose label format. Under Display -> Title, create and edit the map title. The default is simply ATLAS GIS. After you make any final revisions and save them, the maps shown in figures 3-5 are ready to print. (Note that the original maps were printed on a color printer, with different colors representing different ranges of variable values.)

## Conclusion

While the time required to extract, manipulate, and map census data varies with the complexity of the request, much of the process can be automated. The STF outfiles must be manually extracted, but the production of the base map and the joining of outfiles and conversion to a spreadsheet can be done without supervision. Manipulation of data in Lotus 1-2-3 is quickly accomplished, as is the actual production of the thematic map. The maps designed in this tutorial would require between thirty and forty-five minutes to produce.

UNH has been offering specialized data services for almost two years, and so far it has been able to meet the demand for these services. Only recently has the library begun to offer mapping services, but experience

suggests that only a small portion (social scientists, town planners, and economists) of patrons will use this service, and so demand can be fulfilled.

The scenario described above requires the acquisition of software and hardware beyond what depositories

are accustomed to using. It also requires additional training for depository staff. While electronic products present depository libraries with an enormous challenge, such libraries have in these products a wonderful opportunity to fulfill the underlying spirit of the Depository Library Pro-

gram: to put government information in the hands of the public. In a day when the existence of the Depository Library Program is being questioned, services such as providing electronic census products can only strengthen arguments for the program's continuation.

## STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

*Information Technology and Libraries*, Publication No. 07309295, is published quarterly by the Library and Information Technology Association, American Library Association, 50 E. Huron St., Chicago, IL 60611-2795. The editor is Thomas W. Leonhardt, Director of Library Technical Services, Bizzell Library, University of Oklahoma, 401 W. Brooks, Norman, OK 73019-0528. Annual subscription price, \$25.00. Printed in U.S.A. with second-class postage paid at Chicago, Illinois. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 *only*), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

## EXTENT AND NATURE OF CIRCULATION

(Average figures denote the average number of copies printed each issue during the preceding twelve months; actual figures denote actual number of copies of single issue published nearest to filing date: December 1993 issue.) Total number of copies printed: average 7,830; actual 7,630. Sales through dealers, carriers, street vendors and counter sales: none. Mail subscription: average 7,170; actual 7,109. Free distribution: average 108; actual 108. Total distribution: average 7,278; actual 7,217. Office use, leftover, unaccounted, spoiled after printing: average 552; actual 413. Total: average 7,830; actual 7,630.

Statement of Ownership, Management and Circulation (PS Form 3526, Jan. 1991) for 1992/93 filed with the United States Post Office Postmaster in Chicago, September 30, 1993

## MachTen 2.1: UNIX for Macintosh

Tenon Intersystems, 1123 Chapala St., Santa Barbara, CA 93101; (805) 963-6983; fax (805) 962-8202; Internet info@tenon.com. System requirements: Macintosh Classic or better. The reviewer recommends a 68030 processor or accelerator board. At the time of writing, MachTen would not run on a 68040 processor, but Tenon Intersystems stated that an 040-compatible version would be available by the end of 1993. System software: System 6.0.7 or later. RAM: At least 4 MB and at least 5 MB for System 7. Eight megabytes is recommended for using the software-development tools. Disk storage: 29.5 MB to install. After installation MachTen Personal Edition occupies about 26.3 MB of disk space. Price: Personal MachTen \$495; Professional MachTen \$695; optional documentation \$300; X-Window package \$350; academic discounts up to 50 percent available.

Tenon Intersystems has been porting UNIX to Macintosh since 1989. Its most recent releases, the Professional and Personal editions release 2.1, improve on its earlier versions and continue to give owners of relatively low-cost Macintoshes the option of acquiring UNIX capabilities. Not merely a set of UNIX-like tools, MachTen offers Mac users Berkeley Software Distribution version 4.3 of UNIX complete with multitasking and networking capabilities as well as the rich set of utilities and C programming tools that come with Berkeley UNIX. Libraries can use MachTen to set up a UNIX server on Macs as humble as a Classic, at a very reasonable price.

### Two Versions

The basic UNIX package comes in two editions, the Personal and the Professional. The Personal edition is a complete port of 4.3 BSD UNIX to the Macintosh with the utilities that UNIX users would expect on a workstation and some that they might not expect,

such as a style checker and GNU utilities like gzip.<sup>1</sup> Both the C and the Bourne command shells are available, as is the vi text editor. In addition, a good set of C development tools is also included: the GNU C compiler, make, lex, yacc, lint, the GNU debugger gdb, and more. Perhaps the major value to libraries of a C development environment is the ability to build and compile downloaded freeware and shareware for UNIX. Tenon Intersystems maintains an ftp server with instructions on building some of the more popular free UNIX tools on a MachTen system.

The Professional edition has everything in the Personal edition and also a set of software-development tools, including the GNU C++ compiler, which runs only in virtual memory. These tools do not use Macintosh virtual memory available on machines with 030 processors or better, but paged virtual memory, which requires a PMMU chip. Unless a library has a particular need for the C++ language compiler, the Personal edition should be sufficient. The reviewer did not have access to a Mac with a PMMU, so the edition reviewed here is Personal MachTen, tested on a Macintosh LC with a 030 33 Mhz accelerator card and floating point processor, 10 MB of RAM, and an 80 MB hard disk.

### Installation

Installation is quite painless. The six high-density diskettes with compressed files were copied onto the reviewer's hard disk within fifteen minutes. Unpacking the files and installing them took around thirty-five minutes, but the machine did not need to be attended during this period. MachTen must surely rate as one of the most hassle-free UNIX implementations to install. Configuring the installed system will take more or less time depending on the network connections to be established at the site and the number of users to be added.

### UNIX for Most Macs

Tenon states that MachTen will run on a Classic or higher Macintosh. At least 5 MB of RAM are required for running MachTen under System 7; at least 4 MB to run system 6.0.7 or a later version of System 6. To use the software-development tools, at least 8 MB of RAM are recommended. The entire MachTen package will occupy around 26 MB of hard-disk storage. While this may seem like a large bite out of storage, compare it with the 80 MB of storage required for Apple's own version of UNIX, A/UX. The latter also requires a disk partition, while MachTen does not. Once one is familiar with MachTen, it is possible to decrease the disk-storage requirements by deleting features unnecessary to a site's implementation.

### Speed

Though it is technically possible to run MachTen on a Classic, expect the program to run slowly on a 68000 processor. The system was none too nimble on the reviewer's equipment, though speed was certainly acceptable. Tenon makes some good recommendations for system tuning, suggestions that are applicable whether or not one is running MachTen:

Because of the size of the system, obviously not all of MachTen can be loaded into memory, so it must have frequent recourse to the hard disk. To increase hard-disk speed, consider installing a faster disk driver than the one Apple provides. Examples would be those available in the software packages FWB Hard Disk Tool Kit or Silverlining.

Experiment with RAM caching until a setting is found that provides optimal speed. Usually there is a setting beyond which one gains very little in speed and only needlessly takes valuable RAM away from applications. The reviewer found that on his equipment a setting of 128K produced a significant increase in performance over the minimum setting of 32K, but

caches larger than 128K had little additional effect.

## Integration of the Mac OS and UNIX

Since MachTen is an application running on top of the Macintosh operating system, a certain degree of integration between the two systems is possible. One can run Macintosh applications simultaneously with UNIX applications. It is possible to copy and paste among open Macintosh applications and UNIX terminal windows. MachTen can use the native Macintosh Finder as well as the UNIX commands for file management. Text files may be exchanged between Macintosh and UNIX using Tenon's utilities `unixmap` and `mactext`.

MachTen's terminal windows have a Macintosh menu bar. Copy and paste as well as variable fonts are all supported. The windows menu item provides window management capabilities. Figure 1 shows two open terminal windows with documents open

under the `vi` text editor. Anyone who has tried to move text among open files on a UNIX terminal that shows only one window at a time will appreciate the ability to display multiple windows and exchange data among them.

## Networking

The term "Personal" in Personal MachTen might lead one to think that this edition is for single-user workstations only. Personal MachTen, however, supports the same networking options as does the Professional edition: Ethernet, Token Ring, and LocalTalk. File sharing is supported through NSF, the file-sharing standard developed by Sun Microsystems. The program has full facilities for TCP/IP and telnet. A suite of UUCP communications tools is also available, as is software for establishing SLIP connections to the Internet over a serial line. In networking and communications as in hardware requirements, MachTen allows the user to go

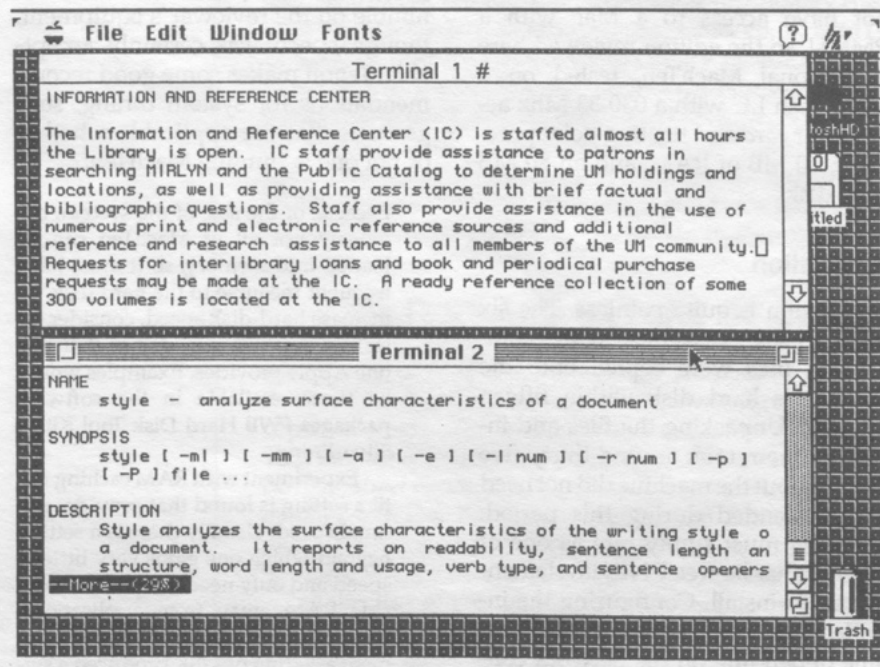
with less expensive options such as SLIP connections and LocalTalk. Libraries with tight budgets should appreciate this flexibility.

## Documentation

The basic package comes with a stapled installation and configuration guide as well as two spiral-bound manuals: *MachTen User's Guide* and the important manual entitled *MachTen System & Network Administration*. The manuals contain classic articles on their subjects by some of the famous names in UNIX history. Tenon has edited some of these texts to take account of MachTen's specific implementation of UNIX.

Additional optional documentation is available for \$300. This set of manuals includes:

- A two-volume reference on the GNU C Compiler
- MachTen UNIX Basics
- MachTen Text Processing
- MachTen Reference Manuals: Two large looseleaf manuals, with entries on all the UNIX commands and a guide to the C libraries. These manuals' contents are also online in the "man" pages, which can be called up by typing "man" followed by the command name on the command line. These same man pages may also be explored using a HyperCard stack called ManTen.
- MachTen Programmer's Reference: These two volumes contain chapters on the various C programming tools available, such as `lex`, `yacc`, `lint`, and so forth. Most libraries will not need the optional documentation. Much of the information contained in these guides is available in other publications, many of which are probably already owned by libraries with computer science collections. The books published by O'Reilly and Associates are especially recom-



**Figure 1**  
Multiple UNIX terminal windows open for the same login.

mended for guidance on a variety of UNIX topics.

## X-Windows

X-Windows, the graphic interface for UNIX, is not included in either the basic Personal or Professional editions. An X package with server and client software and Motif development tools is available separately from Tenon for \$350.

## Conclusion

Personal MachTen can be highly recommended to libraries that want to acquire UNIX capabilities but do not have the funds to spend on expensive UNIX workstations. Libraries that already own and use Macintosh equipment will find the price and power of MachTen especially attractive.—*Brian Sealy, University of Michigan.*

## Note

1. GNU stands for "GNU is not UNIX." GNU is a project of the Free Software Foundation of Cambridge, Massachusetts. The goal of the project is to write a freely distributable operating system and set of utilities and tools compatible with UNIX but using no copyrighted code.

## Microsoft Encarta

Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052-6399; 1-800-426-9400. Hardware and software requirements for playing sound, running animations, and displaying text and graphics: multimedia PC with 386SX or higher processor, 2 MB RAM or more and 30 MB hard disk (an equivalent PC with a multimedia PC upgrade kit, including CD-ROM drive and MPC-compatible audio board, may be substituted), CD-ROM drive, audio board, mouse, VGA or VGA+ display, MS-DOS 3.1 or later (provided with CD-ROM drive), and headphones or speakers. Hardware and software requirements for displaying graphics and text only: PC with 386SX or higher processor with 2 MB RAM, hard disk, CD-ROM drive and VGA

or VGA+ display, MS-DOS 3.1 or later, Microsoft Windows 3.1, MS-DOS CD-ROM Extensions 2.2 or later, and Microsoft Mouse or compatible. Retail price: \$395. Educational price: \$249.

Microsoft Encarta is a multimedia CD-ROM encyclopedia produced by the Microsoft Corporation. It is based on *Funk and Wagnalls New Encyclopedia* and includes a dictionary and thesaurus that are based, respectively, on *Webster's Electronic Dictionary, Concise Edition*, and *Webster's Electronic Thesaurus*.

## Installation

Installation of Encarta is quite easy. Assuming you have a computer already set up with Microsoft's Windows operating system, all you need do is insert Encarta into the CD-ROM drive, open the program manager window, and choose "run" from the file menu. You then enter the drive letter for the CD-ROM drive in the command line box, followed by :\\setup (for example, d:\\setup), and press enter. The Encarta Setup program will create a Microsoft Multimedia group window within the program manager, and you will be ready to begin using Encarta.

## Operation

The opening screen of Encarta provides the user with four choices: to open the encyclopedia, to use the category browser, to view the encyclopedia contents box, or to take a quick tour. Experienced researchers will want to open the encyclopedia or use its contents box, while young researchers will probably find the category browser helpful as they attempt to narrow their area of interest. Category browser is more directive than the other choices, allowing the user to first select an area of interest, such as performing arts, and then see categories within the area of interest, such as music or dance. From there, a user

may choose a narrower topic, such as Fred Astaire or ballet. Category browser options are also included and enhanced through the find button on the main button bar, which is displayed at the top of any Encarta entry. The find button opens the find box, which allows a user to type in a topic and limit where it may be searched by area of interest, time period, biography or nonbiography, and gallery type (i.e., photos, maps, etc.).

Once a user reaches a topic of interest in Encarta, a variety of information is available through the use of the main window, which is the format providing information to the user. The main window displays the category in which one is located; gallery features, such as photos or animation; the text of the entry (called an article by the documentation) under view; and the main button bar, which helps the user navigate through Encarta.

Long articles are moved through with the use of scroll bars, and the display of all articles may be enlarged at will. A frame at the bottom of the area where the text of the article is displayed allows the user to select "see also" references for viewing or to see an outline of the article before scrolling through it. The outline feature is very useful, because by clicking on a heading or subheading in the outline, the user is moved to that part of the article. The outline also indicates where gallery features, such as charts or maps, are located in the article.

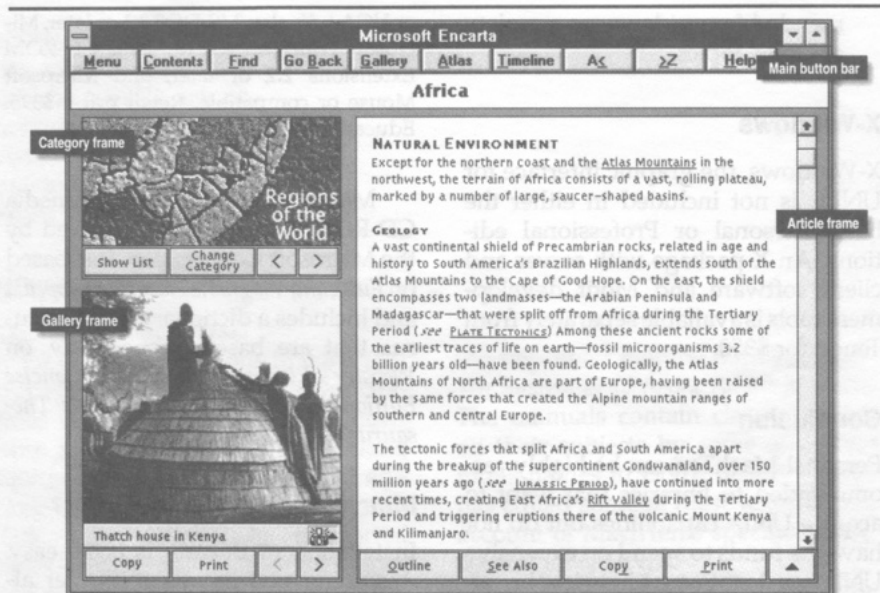
Articles also have colored and underlined words and phrases, which are cross-references. By using the mouse to click on a cross-reference, the user jumps to that topic in Encarta. And when the user is ready to go back to a previous topic—here's a nice feature—he or she clicks the menu from the main button bar, selects "topic tracker" from the menu, and views a list of topics that he or she has visited. The user can then click the topic to which he or she wishes to return.

The gallery is one of the "gee whiz" features of Encarta's window.

The gallery, which takes up about a three-by-two-inch area on a fourteen-inch screen, offers the user audio, photos and illustrations, animations, maps, charts, and tables to enhance articles. Nearly all articles have gallery features, most of which may be enlarged to display greater detail, and gallery items are copyrighted or credited as needed. Major gallery contributors include Dorling Kindersly and Smithsonian Folkways. Examples of gallery items I viewed include animation of the ecliptic (sun's path), a photo of the main reading room in the Library of Congress, an illustration of a scriptorium, a map of Mexico, and an audio clip of a balalaika. Especially impressive are the audio items. I heard a segment of FDR's "a day that will live in infamy" address to Congress, the cry of an American goshawk in flight, and the music of Creedence Clearwater Revival, among others. An aural disappointment was the national anthems of countries, which are synthesized and sound tinny.

The final element of the main window is the main button bar. This "control panel" helps the user select and find new topics, return to previous topics, get help, or browse alphabetically through all Encarta topics. It also provides access to the menu, gallery browser, atlas, and timeline, all of which deserve further attention.

The menu can help the user create electronic notes and bookmarks for future reference, can access the topic tracker mentioned above, and can access an online dictionary and thesaurus. The menu also provides access to a word-processing program that allows the Encarta user to write reports and research papers. The word-processing program is an attractive feature because it allows the pasting of text and graphics from Encarta into a report and automatically includes proper citation information. Educators will be glad to note that Encarta documentation very strongly explains the importance of acknowledging sources.



**Figure 1**  
The four parts of the Encarta Main Window

The gallery browser allows the user to browse through gallery items, locate particular topics in the gallery, see a "slide show" of gallery items, either in alphabetical or random order, or view special lists which provide a range of gallery items on preselected topics. The special lists are terribly fun, with topics such as mammals, in which one can see a bird in flight or hear its call; classical music, with excerpts from the works of famous composers; or famous people, including sound bites from Amelia Earhart, Eleanor Roosevelt, Bill Clinton, and Bill Gates (hmm, wonder how he got in there?!). You can even hear the languages of the world and learn how to count to ten in them or say hello, goodbye, and thanks.

The atlas allows the user to select a continent to explore or type in a place-name or geographic feature and display it in map format. The user can enlarge maps for greater detail, hear pronunciation of names, explore sights and sounds of a region, or move from a map to an appropriate encyclo-

pedia entry to learn more about a topic.

The timeline offers an overview of history, from prehistory to modern times. A scroll box moves the user through the timeline, and all arrows and icons in the timeline are "hot"—that is, the user may click on them to receive more information. Clicking on arrows retrieves information on a time period or culture, and clicking on an icon provides data on an event or displays a gallery item, such as a map.

## Evaluation

It's not hard to see how Encarta could become addictive. The gallery items are the neatest things to hit encyclopedias since color photos, and the entries themselves are designed to encourage learning, leading the user from one entry to another (this is why the topic tracker is such a useful feature; it doesn't take long before the user is far from his or her originally selected topic).

The text of the encyclopedia is not "dummied down," and this is good, as it will encourage readers to increase their vocabulary as they read. For this reason, the online dictionary is very useful. Encarta also provides a number of features designed to stretch users, including the online thesaurus, which encourages users to find new ways to express themselves, and the copyright information included in all pasting activities, which impresses upon users the importance of giving credit where credit is due. The atlas and timeline also provide strong learning opportunities, helping users to place events in proper historical context and improve their geographic skills.

Naturally, Encarta is not a perfect product. Its coverage is less than com-

plete, although it does a good job of not being too U.S.-centric, and it does have its errors, like a graphic icon in a place where there is actually an audio gallery item. Also, I often found myself waiting on a graphic to enlarge or an aural gallery item to begin playing. While this may be as much a CD-drive problem as anything, it could prove to be problematic in an age in which more and more users expect faster and faster results.

### Recommendations

I have no qualms about recommending Encarta for home use, especially for children. Kids will probably spend hours on it, having fun and learning at the same time. I would also recommend this for elementary and junior

high schools, and maybe even high schools and public libraries, depending upon their budgets and clientele, but users would have to be limited in their use time on the product. (Networking would be possible, but would probably noticeably slow down response time, and Microsoft does not support networking this product.) You might also want to provide headphones so that the sound elements do not disturb other library users. If sound is a problem, it is possible to have Encarta display graphics and text only, but what is the fun in that? It would be very frustrating to a user to learn that, even though Encarta provides it, he or she can't hear the sound of a dijeridoo or see an animated demonstration of gravity.—*Karen Rupp-Serrano, University of Oklahoma.*

# GET WIRED

with These New Titles from ALA Books

## Coming This Spring . . .

### **The Internet Troubleshooter: Help for the Logged-On and Lost**

*Nancy Regina John and Edward J. Valauskas*

This compact question-and-answer book is an Internet surfer's lifeguard designed to use right at your computer. The spiral-bound guide is distinguished by its practical approach, sensitivity to common problems, and accessible presentation of answers. Use the index and tabs to flip right to the answers to your user dilemmas. Abundant illustrations and real-life examples make solutions easy to understand and simple to enact.

\$20.00spiral • 1994 • approx. 100p.

**ALA Order Code 0633-8-0011**

## **From the Library and Information Technology Association . . .**

### **Access to Information: Materials, Technologies, and Service for Print-Impaired Readers**

*LITA Monographs, #2*

*Tom McNulty and Dawn Suvino*

Outlines the diverse ways that low-vision readers can access information. Included are a crash course in computer literacy, tactile reading methods and materials, audio output systems, materials and technology, programs and services, and much more.

\$28.00pbk. • 161p. • 1993

**ALA Order Code 7641-7-0011**

### **Telecommunications, Networking and Internet Glossary**

*LITA Monographs, #4*

*George S. Machovec*

Provides hundreds of definitions to telecommunications, networking, and internet terms to help librarians, information managers, and students remain on top of the new terminology in the expanding communications universe.

\$18.00pbk. • 106p. • 1993

**ALA Order Code 7697-2-0011**

## New from ALA Books . . .

### **Guide to Selecting and Acquiring CD-ROMs, Software, and Other Electronic Publications Acquisitions Guidelines #9**

*Stephen Bosch, Patricia Promis, and Chris Sugnet*

Provides all the necessary direction and suggestions for selecting and acquiring collections of electronic publications librarians need. Issues from policy concerns to leasing and rental agreements demonstrate the guide's level of concise detail. New sets of selection considerations not weighed when acquiring traditional formats are included as well as a glossary of commonly used electronic-format terms.

\$10.00pbk. • 1994

**ALA Order Code 0629-X-0011**

### **Guide to Technical Services Resources**

*Peggy Johnson, editor*

This new authoritative guide to more than 2,000 recent print, electronic, and expert sources is essential for every technical services department and professional collection. For the practitioner, it's a first-stop problem solver; for the researcher, an up-to-date bibliography; and, for the educator, a technical services teaching tool.

\$60.00cl. • approx. 350p. • 1994

**ALA Order Code 0624-9-0011**

### **Introduction to Automation for Librarians, Third Edition**

*William Saffady*

Describes information technology as it relates to library operations. Coverage of electronic document imaging systems, programming languages, and prewritten software packages has been greatly expanded. Discussions of bibliographic utilities, CD-ROM cataloging import products, and online public access catalogs have all been completely updated.

\$45.00cl. • approx. 430p. • 1994

**ALA Order Code 0628-1-0011**



ALA Books  
American Library Association  
50 East Huron Street  
Chicago, IL 60611

To order, call 800-545-2433 and press 7



## Buyers and Borrowers: The Application of Consumer Theory to the Study of Library Use

By Charles D. Emery. New York: Haworth, 1993. 188p. \$26.95 (ISBN: 1-56024-183-7).

Initially intended to be a test of the old saw "engineers do not use libraries," this work, happily, grew into much more. Applying consumer theory to library users is a unique and enlightening approach to borrower patterns, and Emery has done an admirable job here. Relying heavily on and skillfully blending the literatures of advertising, statistics, marketing, and information science, this work undertakes a daunting task. Emery identifies eight concepts common to the fields of marketing and library services; for example, "product development." Marketing pros use this to create a total package that will appeal to consumers and make them more apt to buy a product. Libraries do this with the process that we apply to materials to get them ready for our consumers: selection, cataloging, display, and publicity. We, just like marketing managers, want our patrons to use the products we present to them. If they don't, we are also out of business. Many illustrative tables and figures are included. Some of these may be somewhat intimidating for the "statistically challenged" (like this reviewer), but they do point to the validity and thoroughness of Emery's research. Emery introduces each chapter with an appropriate quotation—a nice touch. Of course, a bibliography and index are included. This is an excellent work to be read by students, library administrators, and branch managers, to aid in budget projection and subject area collection. Highly recommended.—*Lisa K. Miller, American Graduate School of International Management Library.*

## CD-ROM Buyer's Guide and Handbook

The Definitive Reference for CD-ROM Buyers, 3d ed. By Paul T. Nicholls. Wilton, Conn.: Eight Bit Books, 1993. 732p. paper, \$44.95 (ISBN 0-910965-08-0).

Three features make this book a valuable reference tool: the reputation and expertise of the author, a thorough treatment of the subject, and the additions and revisions contained in the new edition. Paul T. Nicholls, a professor of library and information science at the University of Western Ontario, has written a plethora of reviews, articles, and books tracing the development of CD-ROM technology and evaluating information-oriented CD-ROM products. Earlier editions of this work, under the title *CD-ROM Collection Builder's Toolkit*, were extensively and favorably reviewed, being labeled "indispensable" and "a must purchase." The third edition is likewise an essential reference that will serve the needs of both neophyte and expert, anyone who needs to know something—or almost everything—about CD-ROM in libraries.

As the introduction indicates, the book's fifteen chapters cover three areas: "background on the CD-ROM medium and industry and practical techniques for buying, using and managing CD-ROM"; "documentary sources of information on CD-ROM"; and "brief, evaluative reviews of about 200 core CD-ROM titles."

Chapters 1-9 provide a wonderful cumulation of all manner of information about the CD-ROM industry and technology. Want the names of CD-ROM-related conferences or a list of reviews of NEC CD-ROM drives? How about an explanation of "multisession" technology? It's all here in text, sidebars, charts, graphs, and bibliographies. Both author and publisher deserve congratulations for

producing a technical overview that's not just readable but also fun to read and easy to use.

Chapter 10 (over 235 pages) offers Nicholls' evaluation of some two hundred "core" CD-ROM titles. The information is presented in film-review style, beginning with essential details about the publisher, the number of discs, hardware requirements, software, price, etc. A text commentary is then followed by a one- to four-star rating of documentation, installation, data quality, search power, and ease of use. Each entry concludes with a brief list of published reviews.

The remaining chapters gather an array of documentary information: an extensive bibliography of CD-ROM product reviews, a classified bibliography of all kinds of other CD-ROM source material, a list of publishers and distributors, and a glossary. The organization and layout of *CD-ROM Buyer's Guide and Handbook* make it easy to manage in spite of its bulk.

Separate lists of tables, figures, and sidebars, and a well-done index, add to the work's utility.

Updated references, much wholly new material, and improvements in organization make the third edition well worth the purchase price. The list of core CD-ROM titles has been expanded from just over one hundred to two hundred items. The directory of publishers and distributors is also nearly double in size. The chapter on CD-ROM source material has new sections on the Internet, Kodak Photo CD, SGML, fonts and clip art, and maps and charts, to name only a few. Multimedia, a minor topic in the second edition, now merits an entire chapter. Even the sidebars hold useful tidbits on such topics as sound boards, licensing, buying a printer, and theses and dissertations on CD-ROM, all new to this edition. As the book's subtitle suggests, this is, as nearly as any one volume can be, a "definitive reference for CD-ROM buyers."—*David Billick, University Microfilms Inc.*

## The Electronic Journal: The Future of Serials- Based Information

Ed. Brian Cook. New York: Haworth, 1993.  
106p. \$24.95 (ISBN 1-56024-452-6).

These seven papers were first presented during a one-day seminar at Bond University in Australia in May 1992 and were subsequently published as volume 3, number 2 of the *Australian & New Zealand Journal of Serials Librarianship*, as well as the monograph under review here. The collection attempts to canvas the broad, rapidly changing horizon of electronic journals (e-journals). All of the contributors—which include librarians, a consultant, a research fellow, and a publisher—work in Australia, but this fact in itself should not devalue the book for readers in the United States.

No one paper rises above the rest. J. L. von Hagen focuses on how intellectual property can be protected in an electronic environment. The biggest obstacle is the absence of pricing models for these projected electronic copyright reimbursement transactions. Von Hagen suggests that the electronic process, both in academic and business environments, has a high profile but low actual use. T. Cochrane emphasizes the balanced triangle between scholar, publisher, and library. From a reference perspective, C. Newton-Smith examines the use of e-journals available on the Internet at no cost. A. H. Newell presents the results of survey research, focussing on CD-ROM technology, conducted at one university. J. Schmidt provides the most complete presentation of the variety of e-journals, then concentrates on full-text CD-ROM products in a state library setting. J. Stevens addresses the development of image databases in a CD-ROM format. At times this chapter reads like an advertisement for a particular product line of image databases. By adopting a skeptical attitude toward e-journals, based on an analysis of the economics of

information transfer, D. Lamberton provides a cautionary note to the book. The editor includes a summative final chapter, akin to a seminar wrap-up.

If the aspects of e-journals include the technological, economic, psychological, political, and legal (e.g., copyright), this book focuses on the economic aspects. Unfortunately, some of the contributors cannot resist the temptation to report how well their organizations are handling this new situation. Most of the contributors refer to the concepts of e-journals, e-serials, and electronic scholarly communication as if they were interchangeable. Although the editor states that the papers have not been edited after the day of the seminar, one of the chapters cites PACS-L messages posted in August 1992, three months after the seminar. Most of the contributors do not address the topic of the book's subtitle, the future of serials-based information, which may be different from the future of serially distributed information. Oddly, none of the contributors addresses the increasingly blurry distinction between monographs and serials.

At \$24.95 and \$19.95 respectively, both the hardback and paperback editions seem too expensive for such a slight volume. In short, this is a disappointing paper-based monograph and a tepid treatment of a hot topic. Wait for the fourth iteration in an electronic version, but don't be disappointed if it never appears. The book is not indexed. Five of the papers include endnotes.—*Thomas A. Peters, Mankato State University.*

## Format Integration and its Effect on Cataloging, Training, and Systems

ALCTS Papers on Library Technical Services and Collections, no. 4. Ed. Karen Coyle. Chicago: American Library Assn., 1993. 109p. paper, \$20; \$18 ALA members (ISBN 0-8389-3432-3).

This publication is based on papers presented at an ALCTS preconference on format integration held during the 1992 ALA Annual Conference. In the introduction, editor Karen Coyle indicates that papers have been altered to make them appropriate for publication. She also writes that the volume "provides an overview of Format Integration and its history, discusses its effect on cataloging and systems, and explores training and documentation" [p. viii].

The first paper is an excellent overview of format integration by Anne L. Highsmith. She describes the need for format integration given technological changes in publishing and the economics of maintaining seven different bibliographic formats. Highsmith then outlines the most significant changes in fields, probable changes in systems, and changes in encoding.

Highsmith's article sets the stage for the subsequent four articles on format integration's effects on cataloging. Glenn Patton, Derry Juneja, and Paul Weiss have prepared brief, but thoughtful reports on the key issues for multimedia materials, monographs, and serials, respectively. There are many very useful illustrations showing cataloging before and after format integration throughout these chapters. Examples are varied and range from the most straightforward monograph to the most complicated multimedia kit to an archivally controlled serial. The examples in Glenn Patton's article, "Multimedia Materials," are particularly striking. The advantages of removing obsolete fields and of merging the various formats to reduce overlapping data, and those created by the new 006 field (additional characteristics of the item) and repeatable 007 fields (Physical Description Fixed Field), are addressed thoroughly.

Other articles appearing in the volume are "Training Issues," "Documentation," "The Utilities," "Local Systems," and "Online Public Access Catalogs." Chapters on systems raise

many issues that utilities, vendors, and libraries will need to consider when determining how to implement format integration. Contributors Jo Calk and Priscilla Caplan indicate that utilities and local systems vendors will have to work with their customer bases to reach answers and that approaches may be very different.

Karen Coyle's closing article on the impact format integration will have on catalog users is very good. Ultimately, the extent to which local systems use the 006 field to display and index multiple formats will determine the significance of the change to an integrated format. As Coyle states, however, "Format Integration represents an evolution of the USMARC formats, not a revolution" [p. 98].

It is assumed that readers are well-acquainted with AACR2 and USMARC codes. To help those who aren't, however, the name of the tag from *USMARC Format for Bibliographic Data* follows the first time a field tag number appears.—*Katherine L. Walter, University of Nebraska-Lincoln.*

## Interfaces: Relationships between Library Technical and Public Services

By Sheila S. Intner, Englewood, Colo.: Libraries Unlimited, 1993. 265p. \$32 (ISBN 1-56308-059-1).

This is a highly readable and thought-provoking book on library technical services in an era of information revolution. Intner has rewritten and updated her columns for *Technicalities* and collected them under one cover. Each of the almost forty chapters addresses a specific topic, reviews arguments, and offers conclusions.

The focus throughout is on how well libraries serve their users: the crux of the relations between technical and public services. Intner pushes for a more explicit recognition of this

common goal, obscured by day-to-day operations and the traditional division of responsibilities. She tackles problems such as the effectiveness of the catalog, cataloging rules, and subject cataloging terminology; library budgeting; treatment of fiction and serials; incorporating electronic databases into library services; bibliographic instruction; education and recruitment of catalogers; and the present "crisis" involving cataloging backlogs.

Intner recognizes that the catalog is an imperfect tool. In seeking ways to correct some of its major shortcomings, she offers some provocative challenges. Should catalogers be held responsible for the failure of users to find what they want—and, if so, what should they do about it? Her solutions include a more proactive role for catalogers in creating indexes and tailoring catalogs to local user populations.

At the same time, Intner is aware of the budgetary limitations on cataloging resources. One intriguing suggestion for dealing with backlogs is to oblige collection development librarians to take account of these limitations, and acquire only what cataloging departments can deal with.

Some of what Intner advocates is already a reality (as she recognizes). Keyword access in online catalogs alleviates the dependence on outdated subject headings; there are experiments going forward in expanding the catalog with tables of contents.

Probably no librarian will agree with everything that Intner says, but it is hard to argue with her main objective of making the library more hospitable to users, and of making use of new technologies to further this end. The book's format makes it easy to pursue related topics; or it can be read as a connected argument. The whole volume would profit from a more comprehensive index, and book preservation issues are not included. Nevertheless, this will be valuable reading for any technical services or administrative librarian.—*Laurel F. Franklin, City College (CUNY) Library.*

## Library and Information Center Management

4th Edition. By Robert D. Stueart and Barbara B. Moran. Englewood, Colo.: Libraries Unlimited, 1993. 402p. paper, \$28.50 (ISBN 1-56308-134-2).

"A decision is a judgment. It is a choice between alternatives. It is rarely a choice between right and wrong. It is at best a choice between 'almost right' and 'probably wrong'" [p. 60]. This quote from Peter F. Drucker is one of many in this textbook. Quotations from management theorists, including historical figures like Frederick Taylor, Henri Fayol, and Max Weber as well as library writers such as Ann E. Prentice and Herbert S. White, are also included. These excerpts give the reader a flavor of the writings of the theorists as well as enhance the overall readability of this well-written text.

Chapters cover basic management themes such as planning, organizing, directing, staffing, and control functions. Lengthy and well-chosen appendixes provide examples of library management documentation. For example, appendix C includes organizational charts from the Boston Public

Library, the Denver Public Library, Duke University Libraries, and the New York State Library. Other appendixes contain samples of mission statements, job descriptions, performance appraisals, budget documents, and ALA policies.

The brief subject treatment of some topics is the only weakness of this volume. Issues involving women and minorities are mentioned only in passing. A complex topic like cybernetic evaluation techniques receives half a page of discussion, which is not sufficient to cover this difficult and somewhat obscure topic. However, in other subject areas the authors have done a fine job. The final chapter on the pace of change in libraries is both informative and challenging.

Each chapter includes numerous bibliographic notes as well as a section

of additional readings. These carefully selected readings would allow an instructor to provide in-depth treatment of any topic covered in the textbook. Indeed, in the preface the authors recommend supplementing classroom work with in-basket exercises, case studies, and other simulation techniques.

This textbook is a valuable introduction to library management. The book is unlikely to be as useful in general library collections outside of schools of library science given the brief treatment of a number of important subjects. Library practitioners may, however, find the appendixes useful in the same way ARL's SPEC kits provide samples of library documents.—*Valerie J. Horton, New Mexico State University Library.*

## **Planning Second Generation Automated Library Systems**

By Edward M. Cortez and Tom Smorch. Westport, Conn.: Greenwood, 1993. 248p. \$47.95 (ISBN 0-313-28361-3).

*Planning Second Generation Automated Library Systems* is aimed at library systems analysts, automated system administrators, information managers, consultants, and advanced students specializing in library automation. The authors say they assume a good familiarity with the functional and technical aspect of systems. Their stated focus is on expanding the reader's knowledge of needs assessment, costing for automated systems, procurement, and hardware/software options.

The authors do not define "second generation." In chapter 1 they say that such systems are known for their telecommunications capability, going beyond the support of local integrated functions and providing access to remote information from one system to another. Elsewhere they say that 1983 was the dividing line between the

generations. They suggest that less expensive storage, LAN open operating systems, and full MARC records are among the characteristics of second-generation systems. Chapter 6 begins with the assumption that most of the readers will be migrating from their first systems to second-generation systems. In the reviewer's experience, most libraries migrating to another system in 1994 are doing so from a system purchased after 1983, thus presumably moving from a second-generation system. One is left with an unsettled feeling: What are "second-generation systems" and why care? The reviewer's experience as a consultant is that "generations" are unimportant to most librarians; what is of interest is expanding functionality to include the loading of journal citation files, interfaces with CD-LANs, linkages with the systems of other libraries, gateways to remote databases, Internet connectivity, EDI-based online ordering and serials claiming, and a host of other functionalities. It is the inability of a hardware platform or applications software to accommodate the additional functionality that causes a library to migrate from one system to another—even if both are "second generation."

That most of the topics in the book are covered elsewhere can be illustrated by summarizing the chapters: Chapter 1 establishes a perspective for the reader, tracing the development of the turnkey marketplace; chapter 2 deals with the planning process, budgeting, staffing, project management, and public relations; chapter 3 summarizes past procurement practices and discusses the procurement of state-of-the-art systems; chapter 4 focuses on hardware and software developments; chapter 5 discusses special automation needs, including CD-ROM, online database searching, scanning technologies, and telefacsimile; chapter 6 deals with the migration of a database from one system to another; chapter 7 discusses evaluating proposals and negotiating con-

tracts; chapters 8 and 9 discuss site preparation, installation, and postinstallation testing and evaluation; chapter 10 treats networking and connectivity issues; and chapter 11 discusses the process of managing change.

There are several appendixes, including a draft interlocal cooperation agreement, an automation project rationale, a library staff automation survey, a list of cost categories, a sample communication diary instrument, an outline for an RFP, an RFI (request for information), hardware photographs, a vendor matrix, a Dynix connectivity model, a negotiating agenda, and several other documents, diagrams, and lists.

There are many lapses in the book. In discussing the history of the marketplace the authors give the capacity of removable disk packs in the late 1970s as 300,000 bytes, rather than 300,000,000 bytes; in chapter 9 their discussion turns away from the "second-generation" system to discussion of implementation of "the automated circulation system."

A librarian in the target audience for the book will be better served by undertaking a literature search to identify books and articles treating migration, linking of systems, and other issues of concern.—*Richard W. Boss, Information Systems Consultants Inc.*

## **Preservation in Libraries: Principles, Strategies and Practices for Librarians**

By Ross Harvey. New Providence, N.J.: Saur, 1993. 269p. \$50 (ISBN 0-86291-632-1).

## **Preservation in Libraries: A Reader**

Ed. Ross Harvey. New Providence, N.J.: Saur, 1993. 483p. \$50 (ISBN: 0-86291-608-9).

In the forward to *Preservation in Libraries: Principles, Strategies and Practices*

for *Librarians*, the general editors state that the author has given a "framework for further reflection on the problems of preservation and for research into their solutions" [p. xv]. Using this framework of ten chapters, Harvey moves from an overview of the problems, causes, and ways of addressing preservation issues to a discussion of the causes of deterioration of library materials. The first two chapters provide background information which is supplemented by readings in the companion volume *Preservation in Libraries: A Reader*. Chapters 3-8 are devoted to discussions of surveying collections, environment, care, handling and education, disaster planning, maintenance, repair and binding, and reformatting. The content of each chapter is supplemented by readings, many of which are referred to in the text and all of which were written within the past decade. Chapters 9 and 10 discuss technological and cooperative strategies and, finally, the development of a preservation program. A twenty-five-page selected bibliography follows.

*Preservation in Libraries: A Reader* is organized in chapters corresponding to the *Principles* volume, and each chapter begins with a short introduction to the topic and the readings. The "Further Reading" bibliography following each chapter provides reference to older and related writings, the majority from the United States but also representing Australia, Canada, New Zealand, and Great Britain.

The author bases his two volumes on four premises: (1) preservation is a management responsibility at the highest level; (2) all collections need a preservation plan; (3) the preservation plan must be adequately funded as part of the ongoing budget; (4) preservation must be the concern of all library staff at every level and a part of all library routines: it is not just a technical specialist matter which takes place in a separate laboratory.

The ten topical sections in each volume support these premises

through Harvey's clearly written, reasoned discussion of the issues. It is clear that the author feels strongly about the preservation of library materials, that he is realistic about the daunting nature of the problem, and that he candidly recognizes that both librarians and management must work together to ensure that long-term access to collections is maintained.

These two volumes, together or separately, provide librarians, managers, students, and interested readers with a wealth of useful information and a framework for understanding and thinking about the problems and challenges of preservation. There are three areas, however, to which Harvey might have devoted more attention: exhibitions and loans, fund-raising for preservation, and state programs, such as the New York State Conservation/Preservation Program. Fund-raising and state programs are perhaps more common in the United States, thus explaining their omission, but issues concerning the preservation and protection of materials identified for exhibition or loan are international in scope. These omissions, however, do not severely detract from the success of the two titles.

Harvey's contribution to the literature of preservation is a valuable resource for preservation managers, concerned librarians, and enlightened policymakers who understand that "preservation is pointless unless it enhances access; and usability of the collections is the ultimate aim of preservation" [p. 9].—*Duane A. Watson, The Research Libraries, The New York Public Library.*

### **Records Management and the Library: Issues and Practices**

By Candy Schwartz and Peter Hernon. Norwood, N.J.: Ablex. 1993. 313p. paper. \$19.95 (ISBN 0-89391-998-5).

*Records Management and the Library* de-

fines records management for information professionals who are required or who wish to take on duties in this field. Written and compiled by two faculty members at the Simmons College Graduate School of Library and Information Science, the text closely follows Simmons' course in records management.

The first part of the text (chapters 1-8) clearly and thoroughly describes standard records management, including an overview and comparison with library science and archives management; historical background; how to apply systems analysis to records management; how to inventory, file, and store records; how to decide what to keep and what to throw away; and how to implement a vital records program. Chapters 9-13 deal with specific issues related to enlarging the role of records management in an organization. Each chapter of the text concludes with a list of references and a set of discussion points that could be used in a classroom or other group setting.

The last three chapters present case studies contributed by Kate Jones-Randall, business reference librarian at the University of Massachusetts, Dartmouth; Nancy Richard and Kathryn Hammond Baker, curator and archivist at the Massachusetts Archives; and Toby Pearlstein, who has held positions as librarian/archivist/records manager at the State Transportation Library of Massachusetts and the Massachusetts Bay Transportation Authority.

The text provides a rich store of helpful information that would allow anyone with a grasp of the basic concepts of library and information science to take on records management duties or to apply records management techniques to any setting. Written with the practitioner or library educator in mind, the information is well organized and presented in not only textual, but also tabular and visual, formats. The chapters include many useful aids for implementing a

records management program, such as sample planning and scheduling charts, sample forms, and lists of records management associations, journals, AV presentations, software directories, etc. There is even a checklist for evaluating database management software.

One of the more interesting charts for the reader with a library background, but little experience with archives or records management, lays out the differences between the three fields. While libraries deal chiefly with organizing materials created by external agencies, records centers focus on internal records created by business operations. Archives can focus on internal records, but the nature of the information is historical rather than functional. These principal differences in sources of materials and nature of information give rise to differences in finding aids and access methods.

This text provides detailed coverage and how-to information for library and information science professionals or educators who need a practical approach to records management.—Karen Calhoun, *OCLC, Online Data Quality Control Section*.

## **Serials Cataloging: Modern Perspectives and International Development**

Ed. Jim E. Cole and James Williams. New York: Haworth, 1992. 415p. \$75 (ISBN 1-56024-281-7) (Also published as *The Serials Librarian*, vol. 22, nos. 1/2 and 3/4, 1992).

Cole and Williams describe this work as an attempt to document the major issues of concern to serials catalogers in the early 1990s. It consists of twenty-three papers divided into four sections: "Education and Training"; "Cataloging Practice, Theory, and Current Developments"; "International Aspects"; and "Options for Change."

"Education and Training" con-

tains two articles. Kathryn Luther Henderson outlines concepts and skills students should learn in a serials unit of a cataloging class and stresses the importance of presenting a body of knowledge as the basis of problem solving, weighing alternatives, and decision making. Jeanne M. K. Boydston reports results of a survey to determine activities for continuing education and staff development among serials catalogers.

"Cataloging Practice, Theory, and Current Developments" contains twelve articles on a wide range of issues, beginning with James W. Williams' overview of serials cataloging from 1985 to 1990. Addressing a major task of serials cataloging, Lori L. Osmus discusses decision making in the recataloging of serials. John R. Riemer considers the value of subject-analysis fields in serial bibliographic records and reports the work of the CONSER Subject and Classification Task Force. Rex Bross, discussing the merits of uniform titles as unique identifiers for serials, suggests that AACR2 would not have worked for serials without this system developed by the Library of Congress. Mitch L. Turitz discusses the presentation of holdings data in union lists and notes that the criteria for selecting a uniform title qualifier, as well as those for determining title changes, have at times contributed to the proliferation of successive-entry records. Adding to the complexity of union lists is the use of a separate bibliographic record for each physical version of a serial publication.

Todd Butler gives an overview of the United States Newspaper Program (USNP) and describes the *Newspaper Cataloging and Union Listing Manual*. Of particular interest is the approach to multiple versions. The "master bibliographic record" concept is based on the paper edition, and all union list holdings, paper and microform, are linked to the master record. Beverly Greer-Butler describes the cataloging and inventory phase of the Maryland Newspaper Project.

Two additional articles address the issue of treatment of multiple versions of titles. John Clark and Wayne Jones describe the cataloging of serial microform reproductions at the National Library of Canada (NLC). The basic NLC practice is to make a separate bibliographic record for each distinct, physical version of a work. As prescribed by AACR2, the microform version forms the basis for the bibliographic description, and details of the original are given in a note. Crystal Graham provides an overview of the microform reproductions and multiple versions issue by describing the AACR2 controversy, Library of Congress (and U.S.) practice, various efforts to study and resolve the problem, and finally describing the two-tiered approach proposed by the CC:DA Multiple Versions Task Force.

The three remaining articles describe copy cataloging of serials at the University of Illinois at Urbana-Champaign; look at the NOTIS OPAC at the University of Iowa as related to serials; and examine serials cataloging time and costs, based on a study at Iowa State University.

The "International Aspects" section contains six articles. Patrick F. Callahan compares the differences and compatibilities between the 1988 revisions of AACR2 and ISBD(S): International Standard for Bibliographic Description for Serials. Albert A. Mullis describes the International Serials Data System (ISDS) and discusses the differences between the ISDS Guidelines, ISBD(S), and AACR2. Three articles describe serials cataloging in the United Kingdom, Australia, and Italy respectively, and the last discusses the status of serials information in current national bibliography.

The "Options for Change" section contains three articles. Mary M. Case and Kevin M. Randall discuss the use of latest-entry cataloging at Northwestern University and present findings of a survey of ARL libraries. They conclude that in some circumstances, latest entry may be preferable to suc-

cessive entry to help users interpret bibliographic records and determine holdings. Mary Ann Sheblé and Carolyn Havens describe the Linked Systems Project (LSP) as it relates to serials. The use of shared authority and bibliographic records could greatly reduce the duplication of effort in serials cataloging by individual libraries. To implement bibliographic record exchange they suggest that LSP move to merge the databases of the three networks and the Library of Congress into a master database (but retain network independence). Lacking that, libraries should at least have the ability to transfer bibliographic records between databases. At a time when the usefulness of the concept of "main entry" in bibliographic records is being questioned, Olivia M. A. Madison makes a strong case for the name main-entry heading in the online environment as a controlled access point and as a formal bibliographic identification for the work, particularly in index and brief catalog displays, to provide adequate information for the user to decide on the usefulness of the records retrieved.

This work collects a wealth of information about serials cataloging in a single volume, providing evidence of the complexities of current issues and the diversity of thought on how to resolve them. I recommended it to serials catalogers and anyone interested in improved access to serials information locally, nationally, or internationally.—Norma Fair, *University of Missouri-Columbia*.

### Other Recent Receipts

*Academic Libraries in Greece: The Present Situation and Future Prospects*. Ed. Dean H. Keller. New York: Haworth, 1993. 226p. \$49.95 (ISBN 1-56024-365-1).

*After the Electronic Revolution, Will You Be The First to Go? Proceedings of the 1992 Association for Library Collections & Technical Services Presidents Program, 29 June 1992*. American Library Asso-

ciation Annual Conference, San Francisco, California: Ed. Arnold Hirshon. Chicago: American Library Assn., 1993. 56p. paper, \$18, \$16.20 ALA members (ISBN 0-8389-7650-6).

Andrews, Christopher. *The Education of A CD-ROM Publisher: A Behind-the-Scenes Tale of Multi-media Intrigue*. Wilton, Conn.: Eight Bit Books, 1993. 222p. paper, \$14.95 (ISBN 0-910965-10-2).

Auld, Lawrence W. S. *Computer Spreadsheets for Library Applications*. 2d ed. Phoenix, Ariz.: Oryx, 1993. 168p. paper, \$47.50 (ISBN 0-89774-724-0).

Basch, Reva. *Secrets of the Super Searchers: The Accumulated Wisdom of 23 of the World's Top Online Searchers*. Wilton, Conn.: Eight Bit Books, 1993. 236p. \$39.95 (ISBN 0-910965-12-9).

Breeding, Marshall. *Essential Guide to the Library IBM PC: Volume 13: PC-Based Integrated Library Systems*. Westport, Conn.: Meckler, 1993. 182p. paper, \$34.95 (ISBN 0-88736-188-9).

Bullen, Andrew H., and Charles H. Davis. *Database Techniques for Librarians: A Primer Using Turbo Pascal*. New York: G.K. Hall, 1993. 207p. \$38.50; paper, \$24.50 (ISBN 0-8161-1968-6).

*Business & Legal CD-ROMS In Print 1993: An International Guide to CD-ROM and CD-Based Multimedia Products in Business, Economics, Finance, Law and Related Fields*. Comp. Regina Rega. Westport, Conn.: Meckler, 1993. 360p. paper, \$55. (ISBN 0-88736-912-X).

Carande, Robert J. *Information Sources for Virtual Reality: A Research Guide*. Westport, Conn.: Greenwood, 1993. 184p. \$49.95 (ISBN 0-313-28804-6).

*CD-ROM Finder: 1993 Edition, The World of CD-ROM Products for Information Seekers*. 5th ed. 1993 (formerly *Optical Publishing Directory*). Ed. James H. Shelton. Medford, N.J.: Learned Information, 1993. 585p. paper, \$69.50 (ISBN 0-938734-70-9).

*Critical Approaches to Information Technology in Librarianship: Foundations and Applications*. Ed. John Buschman. Westport, Conn.: Greenwood, 1993. 248p. paper, \$55 (ISBN 0-313-28415-6).

Daubert, Madeline J. *Financial Management for Small and Medium-Sized Li-*

*braries*. Chicago: American Library Assn., 1993. 270p. paper, \$38, \$34.20 ALA member (ISBN 0-8389-0618-4).

*Democracy and the Public Library: Essays on Fundamental Issues*. Ed. Arthur W. Hafner. Westport, Conn.: Greenwood, 1993. 344p. \$55 (ISBN 0-313-28667-1).

Dewey, Patrick R. *101 Desktop Publishing and Graphic Programs*. Chicago: American Library Assn., 1993. 225p. paper, \$27.50, \$24.75 ALA member (ISBN 0-8389-0606-0).

Elias, Arthur W. *The NFAIS Yearbook of the Information Industry 1993*. Medford, N.J.: Learned Information, 1993. 115p. \$50 (ISBN 0-938734-74-1).

Engst, Adam C. *Internet Starter Kit for Macintosh*. Indianapolis: Hayden Books, 1993. 600p. paper, \$29.95 (ISBN 1-56830-064-6).

*European Directory of Software for Libraries and Information Centres*. Comp. Joanna Wood. Brookfield, Vt.: Ashgate, 1993. 251p. \$69.95 (ISBN 1-85742-092-6).

Gray, Mike, Noel Hodson, and Gil Gordon. *Teleworking Explained*. New York: Wiley, 1993. 284p. \$44.95 (ISBN 0-471-93975-7).

*The International Serials Industry*. Ed. Hazel Woodward and Stella Pilling. Brookfield, Vt.: Gower, 1993. 275p. \$59.95 (ISBN 0-566-07466-4).

*Internet's World on Internet 94: An International Guide to Electronics Journals, Newsletters, Texts, Discussion Lists, and other Resources on the Internet*. Ed. Daniel P. Dern. Westport, Conn.: Meckler, 1993. 451p. paper, \$45 (ISBN 0-88736-929-4).

Jones, Timothy L. *The Americans with Disabilities Act: A Review of Best Practices*. New York: American Management Association, 1993. 120p. paper, \$10 (ISBN 0-8144-2350-7).

Kuhlthau, Carol Collier. *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, N.J.: Ablex, 1993. 199p. paper, \$24.50 (ISBN 1-56750-019-6).

Lancaster, F. W. *If You Want to Evaluate Your Library*. . . . Champaign, Ill.: University of Illinois, Graduate School of Library and Information Science,

1993. 335p. \$39.50 (ISBN 0-87845-091-2).
- Li, Xia, and Nancy B. Crane. *Electronic Style: A Guide to Citing Electronic Information*. Westport, Conn.: Meckler, 1993. paper, \$15 (ISBN 0-88736-909-X).
- Libraries and IT: Working Papers of the Information Technology Sub-committee of the HEFCs' Libraries Review*. Bath, UK: UKOLN, The Office for the Library and Information Networking, 1993. 306p. \$28 (No ISBN).
- Libraries and the Future: Essays on the Library in the Twenty-First Century*. Ed. F. W. Lancaster. Binghamton, N.Y.: Haworth, 1993. \$26.95 195p. (ISBN 1-56024-382-1).
- Losee, Robert M., and Karen A. Worley. *Research and Evaluation for Information Professionals*. New York: Academic, 1993. 240p. \$45 (ISBN 0-12-45770-8).
- Mansfield, Richard. *Desktop Publishing with WordPerfect 6*. Chapel Hill, N.C.: Ventana Press, 1993. 261p. paper, \$24.95 (ISBN 1-56604-049-3).
- Modern Library Technology and Reference Services*. Ed. Samuel T. Huang. New York: Haworth, 1993. 139p. \$24.95 (ISBN 1-56024-458-5).
- Multicultural Acquisitions*. Ed. Karen Parrish and Bill Katz. New York: Haworth, 1993. 322p. \$39.95 (ISBN 1-56024-451-8).
- Multimedia and Networking: Proceedings of the Library of Congress Network Advisory Committee Meeting, December 7-9, 1992*. Washington: Library of Congress, 1993. 56p. paper, \$7.50 (No ISBN).
- Musmann, Klaus. *Technological Innovations in Libraries, 1860-1960: An Anecdotal History*. Westport, Conn.: Greenwood, 1993. 272p. \$55 (ISBN 0-313-28015-0).
- Networking and the Future of Libraries: Proceedings of the UK Office for Library Networking Conference, April 2-5, 1992*. Ed. John W. T. Smith. Westport, Conn.: Meckler, 1993. 220p. paper, \$45 (ISBN 0-88736-863-8).
- New Technologies and Global Restructuring: The Third World at a Crossroads*. Ed. Claes Brundenius and Bo Göransson. Los Angeles: Taylor Graham, 1993. paper, \$55 (ISBN 0-947568-58-1).
- New Technologies and New Directions. Proceedings from the Symposium on Scholarly Communication, the University of Iowa, November 14-16, 1991*. Ed. G. R. Boynton and Sheila D. Creth. Westport, Conn.: Meckler, 1993. 118p. paper, \$35 (ISBN 0-8736-879-4).
- Newby, Gregory B. *Directory of Directories on the Internet: A Guide to Information Sources*. Westport, Conn.: Meckler, 1993. 149p. paper, \$29.50 (ISBN 0-88736-768-2).
- Nolan, Michael J. *PageMaker 5 for Expert Techniques, Macintosh Edition*. Carmel, Ind.: Hayden, 1993. 208p. paper, \$34.95 (ISBN 1-56830-017-4).
- Opening New Doors: Alternative Careers for Librarians*. Ed. Ellis Mount. Washington, D.C.: Special Libraries Association, 1993. 268p. paper, \$36 (ISBN 0-87111-408-9).
- Parker, Roger C. *Looking Good in Print: A Guide to Basic Design for Desktop Publishing*. 3d ed. Chapel Hill, N.C.: Ventana Press, 1993. 423p. paper, \$24.95 (ISBN 1-566-4-047-7).
- Research Libraries—Yesterday, Today and Tomorrow: A Selection of Papers Presented at the International Seminars, Kanazawa Institute of Technology, Library Center Kanazawa, Japan, 1982-1992*. Ed. William J. Welsh. Westport, Conn.: Greenwood, 1993. 504p. \$95 (ISBN 0-313-28966-2).
- Saffady, William. *Optical Disks vs. Micrographics, 1993 ed.* Westport, Conn.: Meckler, 1993. 104p. paper, \$49.50 (ISBN 0-88736-904-9).
- Schneider, Derrick. *The TAO of Applescript: BMUG's Guide to Macintosh Scripting (includes disk containing Applescript)*. Indianapolis, Ind.: Hayden Books, 1993. 236p. paper, \$29.95 (ISBN 1-56830-075-1).
- Security and Crime Prevention in Libraries*. Ed. Michael Chandey and Alan P. MacDougall. Brookfield, Vt.: Ashgate, 1993. 296p. \$59.95 (ISBN 1-85742-014-4).
- Shuman, Bruce. *Cases in Online Search Strategy*. Englewood, Colo.: Libraries Unlimited, 1993. 275p. paper, \$27 (ISBN 1-56308-043-5).
- Tedd, Lucy A. *An Introduction to Computer-Based Library Systems*. New York: Wiley, 1993. 320p. paper, \$41.95 (ISBN 0-471-93788-6).
- The Virtual Library: Visions and Realities*. Ed. Laverna M. Saunders. Westport, Conn.: Meckler, 1993. 165p. paper, \$37.50 (ISBN 0-88736-860-3).
- Walker, Geraldene, and Joseph Janes. *Online Retrieval: A Dialogue of Theory and Practice*. Englewood, Colo.: Meckler, 1993. 380p. paper, \$32.50 (ISBN 1-56308-157-1).
- Wiggins, Bob. *Document Imaging: A Management Guide*. Englewood, Colo.: Meckler, 1994. 187p. paper, \$45 (ISBN 0-88736-869-7).
- Wright, Kieth C. *The Challenge of Technology: Action Strategies for the School Library Media Specialist*. Chicago, Ill.: American Library Assn., 1993. 120p. paper, \$22, \$19.80 ALA members (ISBN 0-8389-0604-4).



# Image.

It's Everything.

## **Putting "Service" into Library Staff Training: A Patron-Centered Guide**

Joanne M. Bessler

Give the patrons what they want. This combination of practical advice, quotes from experts, exercises, and case studies will help libraries create a caring, service-oriented image. Offers concrete suggestions on hiring, training, supervising, and evaluating staff who combine proficiency in their specific jobs with sensitivity to patron needs. A blueprint for libraries wanting to emphasize the service aspect of librarianship.

\$14.00pbk. • Approx. 70p. • 1994 • ALA Order Code 3437-4-0011

## **Guide to Technical Services Resources**

Peggy Johnson

Make your technical services a model of speed and efficiency. This authoritative guide describes more than 2,000 selected print, electronic, and expert sources for modern technical services practice, research, and education. For fast help in problem solving, the guide organizes and annotates the most important sources within technical services in general, acquisitions, descriptive cataloging, classification, subject analysis systems, authority control, filing and indexing, serials management, collection management, preservation, reproduction of library materials, database management, and access services (including circulation).

\$60.00cl. • Approx. 545p. • 1994 • ALA Order Code 0624-9-0011

## **Bare Bones: Young Adult Services Tips for Public Library Generalists**

Mary K. Chelton and James M. Rosinia

This how-to manual for front-line public services staff gives practical advice for libraries serving young adults without a YA librarian. Includes sections on promoting materials popular with YAs, supporting homework needs, working with local schools, and coping with issues related to the age.

\$15.00pbk. • 73p. • 1993 • ALA Order Code 7665-4-0011



**ALA Books and PLA  
For the High-Profile Public Library**

ALA Books • American Library Association • 50 East Huron Street • Chicago, IL 60611

To Order Call 1-800-545-2433, Press 7

# News and Announcements

## Central European Conference and Exhibition for Academic Libraries and Informatics

On September 27–29, 1993, the Central European Conference and Exhibition for Academic Libraries and Informatics was held in Vilnius, the capital of Lithuania, with more than four hundred attendees.

The Lithuanian Librarians' Association joined the International Federation of Library Associations (IFLA) in 1936 and participated in its activities until Lithuania was occupied by the USSR in 1940. For fifty years international contacts of Lithuanian librarians were severed. The Central European Conference and Exhibition for Academic Libraries and Informatics was the first international event in the history of the Baltic library community. Two hundred and ninety-two participants in the event were from Lithuania, twenty from Latvia, thirteen from Estonia, and seventy-six from other countries.

The idea for the conference and exhibition occurred to Lee G. Burchinal (U.S.). He proposed to hold the meeting as one of the means to promote a joint project of four Lithuanian academic libraries on implementing new technologies. Dal Lukas of the National Library of Medicine (U.S.) supported the idea. The Lithuania Technical Library, the Vilnius Technical University Library, Kaunas University of Technology Library, and Kaunas Vytautas Magnus University Library organized the international meeting amid the current economic crises in the Baltics.

The conference and exhibition were supported by the Ministry of Communications and Informatics of Lithuania, especially by Minister Professor D. hab. Gintautas Zintelis and Vice-minister Romualdas Krukauskas.

They offered the facilities of the ministry for the conference. The significance of the event was recognized by the international and national organizations that agreed to cosponsor it: UNESCO, IFLA, FID (International Federation for Information and Documentation), VITA (Volunteers in Technical Assistance), SIGCAT (Special Interest Group on CD-ROM Applications and Technology), and the Federal Government Information Processing Councils.

The slogan of the Central European Conference and Exhibition for Academic Libraries and Informatics was "Empowering users in the twenty-first century." The main theme centered on the premise that the patrons of academic libraries constitute the core of the countries' present and future intellectual power. The economic, educational, and information technology development of the countries depends largely on this core group. Academic libraries of Central and Eastern Europe may well become the centers for promotion of new technologies in the information infrastructure. Academic libraries implementing new technologies today will produce thousands of empowered users in the twenty-first century.

At the opening session greetings were offered by the Minister of Communications and Informatics, Republic of Lithuania; the rectors of the universities whose libraries were involved; and representatives of the Lithuanian parliament and the U.S. embassy. Lee G. Burchinal commented on the main theme of the conference. Romualdas Krukauskas spoke on the Lithuanian state information network, and Abdelaziz Abid discussed UNESCO's General Information Programme and its implications for Eastern and Central Europe. Unfortunately, Ivor Kemp from International Book Development, Ltd., London, was stricken ill in Moscow and could not deliver his paper. He intended to speak on the case for investment in higher education li-

brary and information services in Central and Eastern Europe.

The general session consisted of three major papers on the U.S. academic libraries' philosophy and services. Leading specialists—Richard Fitzsimmons from Pennsylvania State University, Doris R. Brown from DePaul University, and Barbara Lockett from Rensselaer Polytechnic University—described their visions of library services and information technology in the twenty-first century. Fitzsimmons discussed the use of technology in addressing the service, collection, and administrative philosophies of academic libraries. He highlighted how the results of intellectual and research endeavors are communicated by referencing the Pennsylvania State University's Library Information Access System (LIAS). Brown presented the need for careful long- and short-range strategic planning in a consortial organization, related to the development and operation of ILLINET Online. ILLINET Online is the statewide resource-sharing catalog for all Illinois libraries and serves as the online catalog and circulation system for forty-two academic libraries in Illinois. Lockett described the role of the twenty-first century librarian as information handler, designer of access systems, and evaluator of user needs. She illustrated these concepts by describing Rensselaer librarians' work in the design of its Infotrax system.

Breakout sessions over the next two days were divided among attendees' interests. Presenters discussed digitizing, interactive multimedia technology, the Internet, the NLM locator, the Eisenhower National Clearinghouse, and LIBNET. Special attention focused on agricultural and medical information. Papers from Australia, Estonia, Finland, Israel, Latvia, Lithuania, the Netherlands, Norway, Poland, Russia, Slovakia, the United Kingdom, and the United States were delivered.

The exhibition was held in the

Lithuanian Technical Library, a building of the seventeenth century. It was distinguishedly qualified. Automated library systems, hardware, software, and networks for information technologies, subscription services, books, and CD-ROMs were presented. Exhibitors included ALISE (Latvia), Alstakoma (Lithuania), Baltijos konsultacija (Lithuania), Center of Bibliography and Book Science (Lithuania), Elsevier Science Publishers (Netherlands), INFO-VITA (Lithuania), and INTA (Lithuania). Additionally, there were K. G. Saur Verlag (Germany), Bowker-Saur (U.K.), R. R. Bowker (U.S.), K. T. Datacenter (Finland), Russian National Public Library for Science and Technology (Russia), Swets and Zeitlinger (Germany), UMI (U.S.), and VTLS (U.S.).

Special demonstrations were conducted by ADONIS (Netherlands), CAB International (U.K.), CD-ROM, Inc. (U.S.), DIALOG Europe (U.K.), EBSCO (Netherlands), IGLA (U.K.), Knowledge Access International (U.S.), NISC (U.S.), and PAIS (U.S.). Additionally, demonstrations were provided by PsycINFO (U.S.), Sociological Abstracts, Inc. (U.S.), H. W. Wilson Company (U.S.), Winnebago Software Co. (U.S.), and World Book International (U.S.).

Acquaintance with high-quality reports forecasting the future of librarianship as well as with new information technology products gave Central and Eastern European librarians ideas on perspectives of their work today and in the twenty-first century. For Lithuanian librarians, the conference and exhibition were breakthroughs in understanding the role and possibilities of new technologies in library activities. The Lithuanian library community appreciated the goodwill of the exhibitors and vendors who donated their equipment and materials for future use.

Personal contacts made during the meeting have opened wide opportunities in cooperation for the librarians

of East and West. Discussions on how to continue the tradition established in Vilnius are under way. The participants, exhibitors, and vendors have expressed a wish to continue the Central European Conference and Exhibition for Academic Libraries and Informatics as an annual event.—*Richard Fitzsimmons, Director of the Library, Pennsylvania State University, Worthington Scranton campus; Vida Maceviciene, Chief, Automation and Information Department, Vilnius Technical University Library*

## NCLIS Plans Survey of Public Libraries and Internet

The U.S. National Commission on Libraries and Information Science (NCLIS) has announced plans to survey public library involvement with Internet, the global computer network of networks. A quick-response sample survey will be conducted in early January 1994 to obtain baseline data regarding the impact of the Internet on public library services and activities. Survey findings about the extent of public library involvement with and use of network services will be analyzed to identify potential federal policies relating to the public library role in developing a networked information infrastructure for the nation.

NCLIS will publish a project report in early spring 1994. Charles R. McClure, professor at the School of Information Studies at Syracuse University, and Douglas L. Zeizig, professor at the School of Library and Information Studies at the University of Wisconsin, Madison, are coprincipal investigators for NCLIS for the public libraries and the Internet project. Public libraries in the sample survey will be selected from a university file developed by the National Center for

Education Statistics (NCES) in cooperation with the national commission and state library agencies.

Timely, reliable data describing current public library Internet/network use and the impact of the Internet on libraries and the communities they serve are required to implement recommendations made at the 1991 White House Conference on Library and Information Services (WHCLIS) that call for the development of a national "information superhighway" network to "be available in all libraries and other information repositories at every level."

The NCLIS survey project also responds to a question raised at a July 1993 policy discussion meeting held at the Library of Congress. At this meeting, Vice-president Gore asked to what degree public libraries can serve as a "safety net" for access to Internet information and services. The vice-president also raised issues related to how the Internet could provide better social equity for the public. Several policy proposals that affect the development of libraries in this new networked environment are under consideration by Congress and the administration. These proposals require current national data about public libraries and the Internet to make informed decisions about future government roles involving public libraries and the information infrastructure.

The following representative topics and issues illustrate the type of information that is expected from the project:

- Degree to which public libraries have operational connections to the Internet
- Type of provider that the library uses to obtain connectivity
- Number and type of people within the library that have Internet addresses
- Internet services and resources that are used by the librarians and their assessment of these services and resources

- Library programs or services that have been developed that incorporate Internet use
- Barriers and problems related to using the Internet
- Estimated expenditures and costs for connecting to and using the Internet
- Special arrangements by which libraries are connected to or using the Internet (e.g., state network users, federal grant recipients, subsidized Internet access rates, etc.)
- Libraries that are not presently connected but that are planning for Internet access in the near future

The U.S. National Commission on Libraries and Information Science is a permanent, independent agency of the federal government charged with advising both Congress and the president on matters relating to national and international library and information services, policies, and plans.

## RLIN Records Task Force: Institutional Perspectives on RLIN—Summary of Survey Responses

As part of its charge to make recommendations on the form and content of the RLIN bibliographic files and ways to expedite record contributions, the RLIN Records Task Force distributed a brief survey to all RLG member representatives and the directors of (nonmember) institutions that use RLIN for technical processing. The purpose of the survey was to solicit general views about the current and potential functions and utility of RLIN as a centralized bibliographic database in an increasingly decentralized technical environment. A total of 210 institutional users of RLIN were

sent the survey and asked to respond—one response per institution. Eighty-eight institutions responded, for a cumulative response rate of 42 percent. Sixteen of the twenty largest RLG member institutions contributed a response, and the overall response pattern was fairly representative of the overall RLG membership. The following analysis summarizes the responses to the nine-question survey. Attached are a list of respondents and a list of responses, grouped by question and thereunder by type of institutional member (general member, special member, nonmember).

### General Impression

1. All data elements in bibliographic records are important and RLIN records should contain as many of them as possible.
2. There are no significant barriers to record contribution.
3. Many members are not willing or able to contribute many kinds of data they judge to be important.

The reasons for these apparently contradictory findings emerge on closer inspection.

**Question 1.** Please rate the importance of the following features of the RLIN bibliographic files and RLIN system functions to your institution on a scale of 1 (not important) to 4 (very important).

The features of RLIN are listed in order of preference, followed by the average rating given the feature by the entire population of respondents:

- (a) union database (3.7)
- (c) resource for cataloging (3.6)
- (f) vehicle for ILL (3.12)
- (i1) general technical processing system (3.02)
- (i3) technical processing system for archival and manuscript collections (2.52)
- (b) data resource for acquisitions information (2.51)

- (h) data resource for end users (2.41)
- (d) source for local collection building (2.35)
- (g) cooperative collection development (2.27)
- (e) record of preserved materials (2.16)
- (i2) materials written in JACKPHY (2.07)

These composite scores are not enough to evaluate preferences fully. Acquisitions (b), collection development (d), resource sharing and cooperative collection development (g), and end-user data resource (h) all score in the mid-2s because a lot of people rate their importance as 2 or 3. Preservation information (e), JACKPHY (i2), and AMC (i3) score about the same, but with a lot more 1s and 4s. That is, these specialized areas are either very important or unimportant depending upon the type of library.

There were some marked differences between the responses of general members and those of special members and nonmembers, where the use of RLIN for technical processing of ALL materials (ranked last by general members) was rated above ILL by special members and nonmembers. RLIN as a data resource for end users ranked relatively low, even among institutions that have subscribed to Eureka.

Because the survey was distributed before Eureka went into production, it may still be too new to have had an impact on survey responses.

**Question 2.** What additional information needs to be represented in the RLIN bibliographic files to make it more useful to your institution to support the features and functions you view as most important?

**Question 3.** Of the types of information you identified in #2, which ones would your institution be willing to contribute for loading into RLIN? Please answer with the letter(s) a-i, as listed above.

Out of the total number of libraries

**Table 1**  
General RLG Members' Perceptions of Usefulness and Willingness to Supply Data

	USEFUL		WILL SUPPLY	
	%	No.	%	No.
a. Order records	34	31	18	16
b. In-process records	50	45	26	23
c. Volume holdings	70	63	33	30
d. Issue holdings	39	35	14	13
e. Publication patterns	54	49	19	17
f. Extant items	54	49	28	23

**Table 2**  
General RLG Members' Information Needs and Willingness to Contribute

	Y/Y (%)	Y/N (%)	Y/? (%)	N/N (%)
a. Order records	29.6	11.0	3.7	44.4
b. In-process records	33.3	18.5	7.4	37.0
c. Volume holdings	55.5	18.5	14.8	11.1
d. Issue holdings	11.1	14.8	11.1	55.5
e. Publication patterns	22.2	25.9	11.1	40.7
f. Extant items	25.9	18.5	7.4	40.7
g. Contents	40.7	33.3	11.1	7.4
h. Preservation actions	44.4	3.7	18.5	22.2

responding, the percentage and number of respondents that claim to value the information are compared with the percentage and number of respondents willing to supply the data to RLIN (see table 1).

Perceived usefulness of all types of information is significantly higher than willingness to contribute that data. Volume holdings was the only category that had more yes than no on the willingness to contribute. Nevertheless, there is a fairly strong correlation between perceived usefulness and willingness to contribute. A major exception is publication pattern data, which is useful, but which few institutions are creating (so they cannot contribute it).

Respondents ranked the need for table-of-contents data high, next after volume-level information for serials holdings. Issue-level data ranked much lower than volume-level information, both for need and for willingness to contribute. This may be because local systems cannot provide the information. Many institutions voiced the need for "extant item" records retained when the last holding library deletes their record.

Answers to question 7 below suggest that general members are most likely to have local systems in place with the capability of producing bibliographic records for the RLIN database. The question therefore arises about the types of records that general

members find important and may be willing to contribute to the RLIN database. Patterns of consistency between need and willingness to contribute are discernable when answers to questions 2 and 3 are combined.

Table 2 displays the combination answers. The first column is the response category. Column 2 is the proportion of general members with both the need and the willingness to contribute. Column 3 is the proportion of members with the need for information who won't contribute. Column 4 is members with the need who did not say whether or not they would contribute. Column 5 is general members without the need for information who also have no willingness to contribute.

More than half of the RLG general members have a need for and are willing to contribute volume holdings for serials titles; more than a third are willing to supply records for materials received but not yet cataloged, as well as contents notes and information on preservation actions anticipated (queuing records). At the other end of the continuum, there is a strong absence of need for issue information for serials holdings and order records—as well as an equally strong reluctance to supply this information. If one interprets the "Y/?" respondents as possible contributors, the importance of volume holdings increases significantly, as do preservation actions and contents notes. Although a majority of the respondents consider publication patterns useful, few were willing to contribute. The question of whether RLIN should retain a record even when the last record in a cluster has been deleted by the owning institution, considered useful by a majority of the respondents, depends first upon whether institutions regularly send updates to RLIN.

**Question 4.** What information that is currently in the RLIN bibliographic files does your institution think is not needed?

**Table 3**  
Value to Members of RLIN Information

	AVERAGE RATINGS (1-4)			
	GENERAL	SPECIAL	NONMEMBERS	ALL
a. bib. descriptions	3.85	3.88	3.96	3.90
b. citation files	2.81	2.62	2.12	2.53
c. table of contents	3.04	2.55	2.15	2.58
d. full-text files	2.96	2.65	2.04	2.55
e. graphic images	2.78	2.48	1.64	2.31
f. numeric data	2.67	1.62	1.54	1.92

**Table 4**  
Local Systems in Use

	GENERAL	SPECIAL	NONMEMBERS	TOTAL	
NOTIS	15	1	0	16	19%
Innovative	3	5	4	12	14%
Gear	3	0	0	3	4%
None	3	14	10	27	33%
Other*	4	13	8	25	30%
				83	100%

\*Variety of systems

Almost two-thirds of the respondents (fifty-five) did not answer this question. One-quarter of the respondents (twenty-three) specifically noted that none of the information in the RLIN database or the bibliographic files should be deleted. Of the twelve institutions that made specific suggestions, two noted that RLG could delete acquisitions records that have not been updated, and four suggested that multiple records for a single entity could be deleted.

**Question 5.** Please rate the significance of the following possible barriers to contributing records to the RLIN database on a scale of 1 (not significant) to 4 (very significant).

No single barrier had an average rating as high as 3.0 for any category of member. Of the four possible response categories specified in the survey, "different reporting requirements to multiple databases" rated highest (2.81) for general members, who also were more likely to specify this barrier as "very significant." Overall, respondents ranked "costs of generating data for export" as the most significant barrier, although, again, few general and special members rated this barrier as "very significant." Nonmembers, however, singled out cost issues as far and away the most significant barrier to the submission of bibliographic records to RLIN.

Perhaps more important and indicative of respondent perceptions is that only a handful of libraries said that no single barrier was significant. In other words, *any* significant barrier can prevent a library from sending records, even if the barriers are different in different places. Thus, trying to remove any particular barrier probably will not improve the overall situation all that much. There may be no major barriers to contributing records per se, and yet it may be difficult or impossible for institutions to contribute particular types of information because of specific limitations in local systems. The local systems survey may point to clusters of institutions with local systems that have the technical capability to submit appropriate records in appropriate formats. RLG needs to work not only on reducing barriers but also on raising institutional commitment to building the RLIN database.

**Question 6.** Please rate the value to your institution of having the following types of information in the RLIN database on a scale of 1 (not useful) to 4 (very useful) (see table 3).

There is a fairly clear and consistent pattern in the responses to this question. Overall, all types of RLG members view the primary function of the RLIN database as providing bibliographic descriptions of materi-

als in all formats. There is no type of information that ranks a close second. All categories of members gave a relatively high ranking to the inclusion of table-of-contents files in the RLIN database. Nonmembers rated this information second in importance after bibliographic data.

The task force discussed the possibility of including pointers in the RLIN database to large and complex files of image data, numeric data, and full-text databases, rather than recommending that such files be included in RLIN proper. Overall, respondents rated the inclusion of full-text files as third in importance, after bibliographic and table-of-contents data. Yet a significant portion (over one-third) of the general and special members of RLG rated the inclusion of full-text files in the RLIN database as "very useful." Special members rated citation files as more important than table-of-contents files.

**Question 7.** What local system(s) do you currently use at your institution?

It is noteworthy that fully a third of the respondents reported having no local system. Those who do not have local systems are, nevertheless, contributing to RLIN by entering records directly online. Obviously, for these institutions it is impossible to submit records to RLIN on tape or by FTP. The lack of local system capability is clus-

**Table 5**  
Anticipate Changes to Local System

	GENERAL	SPECIAL	NONMEMBERS	TOTAL	
Yes	9	13	5	27	31%
No or N/A	16	20	20	56	64%
Maybe	3	1	1	5	5%
Total	28	34	26	88	100%

tered in the special and nonmember categories.

**Question 8.** Do you anticipate changing to a different local system within the next three years (see table 5)?

The fact that 31 percent of the respondents expect to change local systems within the next three years is significant in that it presents RLG with the opportunity to build closer linkages at the point of transition to a new system.

**Question 9.** What system(s) or resource(s) do you use as a bibliographic data resource other than RLIN?

OCLC is a major bibliographic data resource for many of the respondents, particularly general members. Twenty-five of twenty-eight general member respondents use OCLC in some form. All twenty-five are major academic research libraries. The three that do not use OCLC are small, independent libraries. Eleven of thirty-four special members listed OCLC as a resource. Ten of twenty-six nonmember affiliates listed OCLC.

Beyond this finding, the results of this question are not consistent. Some respondents specifically note that the LC resource file, mounted locally, is an important source of bibliographic information, while other institutions that are known to the task force to have this resource available did not list it in the questionnaire. Additional follow-up will be required if clarification of this

issue is needed by the task force to formulate its recommendations.

### General Conclusions

We can set goals and provide incentives for various types of contributions, but any real "requirements" will have to be modest and coupled with high incentives.

The contribution of volume-hold-

**Table 1**  
From RLIN Profiling Runs Completed 9/28/93

FORMAT	RECORDS	TITLES	REC/CLUSTER
<b>Clustered Files</b>			
BKS	53,317,307	17,363,092	3.07
MAP	246,770	214,455	1.15
REC	1,393,068	990,088	1.41
SCO	1,118,226	608,541	1.83
SER	3,757,839	2,161,938	1.73
VIM	195,356	170,206	1.15
Total clustered files	60,028,566	21,508,320	2.79
<b>Nonclustered BIB files (9/28/93)</b>			
AMC	386,327	386,327	1.00
MDF	34,502	34,502	1.00
Total BIB files	60,449,395	21,929,149	2.76

ings information is a strong member preference, but it may not be easy to measure and monitor compliance. For the near future, we may see an even greater dichotomy between the needs and contributions of classes of users. There is some hope of actively working with NOTIS sites to get much of the information we want, especially if the incentives are there.

## RLIN Record Task Force November 5, 1993: RLIN Titles

The Research Libraries Group has just completed its semiannual RLIN database profiles. One of the most common questions we're asked is "But how many titles [not records] are in RLIN?" Table 1, below, provided by Karen Bendorf of RLG, shows the title and record count for each of the RLIN

bibliographic files and also indicates the files with the highest percentages of unique records.

Your contributions to the RLIN database—whether directly on RLIN or by submitting tapes to be loaded—continue to make RLIN the valuable resource it is.—*Karen Smith-Yoshimura, RLG Library and Bibliographic Service*

## OCLC Users Council Discusses "Bibliographic Commons"

The OCLC Users Council met October 10–12, 1993, to discuss a topic of common interest to those who contribute to and use the OCLC Online Union Catalog—the Bibliographic Commons.

"The Bibliographic Commons and Beyond: Electronic Publishing and Knowledge Management" is the theme for the 1993–94 Users Council. "Preserving the Bibliographic Commons" was the focus of the fall meeting.

Gerald R. Lowell, director of libraries at the University of California, San Diego, compared the historical role of traditional English commons with the contemporary role of the Bibliographic Commons—the Online Union Catalog. He drew parallels from the fate of sixteenth-century commons to areas he saw as potential threats to the Bibliographic Commons.

For example, Lowell compared local systems and regional networks with larger farms. Farmers with forty acres or more found they could graze their cattle without using the common lands. Many libraries now have the equivalent of that forty acres in their local systems. "We need to view local systems and the Online Union Catalog as equally important parts of an

even more important whole—that of the national information network," said Lowell.

Lowell also cited differences in goals and objectives, data and record exchange obstacles, ownership roles and responsibilities, fiscal issues, and the electronic arena as potential threats to the Bibliographic Commons.

"We must support appropriate amounts of research and development, risk taking, and creative entrepreneurial visioning" to combat these threats, he said.

"OCLC's history lies in cooperation and resource sharing, and its future depends on it," said K. Wayne Smith, OCLC president and chief executive officer, in remarks to the Users Council following an update of OCLC operations.

Smith said that in the past twenty years, OCLC member libraries have created a commons—the Online Union Catalog—that they have used and shared for the benefit of their institutions and their users. OCLC, in turn, has provided libraries with more than one billion dollars in products and services that have saved them literally billions of dollars in potential duplicate costs.

"We must recognize that the Online Union Catalog is one very important component among many important components of the commons," Smith added. "For what libraries and OCLC have been building is not just a cataloging commons, but an information commons that includes resource sharing, reference, electronic publishing, and electronic delivery. We simply must continue to grow and preserve this commons for future generations."

Arnold Hirshon, Wright State University librarian, spoke about the future of library technical services. He said technical services operations are shrinking nationwide due to budget cuts and other changes. "We're going to have to start creating a new future for technical services if there's going to be one," he said.

Hirshon said adjustments to meet changing information needs should include focusing on customer perspective, improving productivity, reengineering the processes instead of the tasks, eliminating middle-management bureaucracy, rotating job assignments, maximizing library resources, and hiring multidimensional individuals and providing them with ongoing training.

A four-member panel of Users Council delegates representing different types of libraries responded to Hirshon's thought-provoking presentation.

OCLC staff members made presentations to the Users Council. Martin Dillon, OCLC library resources management division director, introduced topics for interest group discussion on the pricing structure for general member tape loading and OCLC's near-term cataloging strategy. Rick Schwieterman, OCLC vice-president of member services, gave a brief update on new OCLC products and services.

In other matters, the Users Council:

- Heard a presentation by Larry Learn, OCLC technical planning director, on past, present, and evolving telecommunications paradigms
- Heard a report from Bill Crowe, dean of libraries at the University of Kansas and chair of the Code of Responsible Use Task Force, following the first meeting of the task force (The task force will consider the current and evolving ways records from the Online Union Catalog are used for more than cataloging in an increasingly networked environment. A survey will be sent to Users Council and other OCLC advisory groups for input before the December 3 task force meeting in Chicago.)
- Elected Jon Lazar, supervisor of technical services for the Rochester Public Library, to chair the nominating committee (Lazar re-



---

mindful delegates that the Users Council will be voting on two open board positions at the spring meeting. The committee will nominate candidates in January 1994.)

- Unanimously approved the following motion: "OCLC, alone, and in conjunction, as appropriate, with its members, adopt a leadership role in responding to initiatives regarding access to, and organization of, information in the emerging national information infrastructure."

---

## Mead Data Central, Inc., to Support New High-Speed Data Transmission

Mead Data Central, Inc., has announced a new high-speed transmission capability for its customers who use the LEXIS/NEXIS online information services. The company has installed modems that support 38.4 kilobit-per-second data speed.

To download documents from LEXIS, for example, to a user's workstation at 9600 baud might take 2.5 minutes. At 38.4 kilobaud it would take only 45 seconds. To scroll up a full screen of text at the slower speed takes about 3.5 seconds, but less than a second at the faster speed.

"Mead Data Central is the first online service to offer this new capability," said Sam Kapoor, senior director of telecommunications. "We are testing the 38.4 Kbps service now so that we will be able to support this new standard of data speed when the faster modems become generally available for sale sometime in 1994."

The equipment, called VFast modems, has been installed in both law firms and corporate customers of the LEXIS/NEXIS services.

## Eight RLG Institutions Begin Work on Innovative Digital Image Access Project

The Research Libraries Group has launched a one-year collaborative project to explore the capabilities of digital image technology for managing access to photographic collections. Eight RLG institutions will work together to find ways to streamline indexing methods and capitalize on the online digital environment for improving access to collections, not just for local projects but for shared access in a networked environment. In total, eight thousand photographs from the collections of these RLG members will be digitized, and image access software will be created.

### Development, Testing, and Evaluation of System

Working with Stokes Imaging Services (Austin, Texas), the RLG cooperative project will employ state-of-the-art digital-imaging technology to enable institutions to reproduce images quickly and inexpensively at resolutions adequate for browsing, researching, and printing reference copies. The project will develop, test, and evaluate a digital image access system consisting of databases, image bases, and software for searching, retrieving, and displaying images.

Each institution will select one thousand photographs that fit the general theme of "the urban landscape"—a topic broad enough to encompass a wide variety of images, yet specific enough to test integrated retrieval of related but separately housed and cataloged image collections.

The images will be first reproduced as 35mm intermediates and then digitized by Stokes Imaging

Services. Access software will be developed by Stokes Imaging Services in consultation with the project participants, who will also evaluate the system at every stage and test its use as a reference resource and technical services tool.

### A Second, Parallel Project

The project was designed by the RLG Task Force on Photograph Preservation, a six-member group formed in July 1992 to address issues inherent in accessing and preserving large photographic collections. The task force has also developed a second, parallel project, in partnership with Stokes Imaging Services and the Image Permanence Institute, to explore relative cost-benefit trade-offs in the technical aspects of digitization, including image capture, resolution, compression, storage, and display.

The participants in the RLG Digital Image Access Project are the Amon Carter Museum, Columbia University, Duke University, the Getty Center for the History of Art and the Humanities, the Frances Loeb Library, Harvard University, the New York Public Library, Northwestern University, and the University of California at Berkeley.

---

## The Power of Learning with Multimedia Personal Computing

To help educators explain how interactive technology can help the learning process, the Computer Learning Foundation, a nonprofit educational foundation dedicated to increasing children's computer literacy, and Microsoft developed a booklet: *The Power of Learning with Multimedia Per-*

sonal Computing. This booklet does the following:

- Introduces multimedia personal computing and explains how using computers at home can sharpen key learning skills
- Provides simple, fun multimedia computing activities and projects for the entire family
- Outlines the four basic components needed for a multimedia personal computer system
- Includes an easy-to-understand glossary of basic computer and multimedia terms

Single copies are available free to educators and parents by calling 1-800-426-9400. Educators can order multiple copies in quantities of fifty for \$11 by sending a check to the Com-

puter Learning Foundation, Attn: Multimedia Booklet, P.O. Box 60967, Palo Alto, CA 94306.

## OCLC Issues Cataloging and Database Services Strategy

OCLC's *Cataloging and Database Services Strategy* has been sent to the OCLC membership. A fact sheet on the OCLC Online Union Catalog is included with the White Paper.

The four-page document outlines OCLC's cataloging enhancement strategy through the remainder of the decade.

The paper sets forth a new vision for cataloging as libraries move further into the electronic age. The strategy includes the following:

- Improving productivity by further automating the copy cataloging and authority control processes
- Facilitating links between local library systems and OCLC systems and services to increase library access
- Providing authority control services
- Enriching the OLUC to support the needs of scholarship

To obtain a copy of OCLC's *Cataloging and Database Services Strategy*, write to Cataloging White Paper, OCLC, MC 204, 6565 Frantz Rd., Dublin, OH 43017-3395.

# Letter

To the Editor:

Aren't information scientists grand as they, like other true scientists, investigate the important issues of Our Profession? I was initially intrigued by—but ignored—the communication from Terry Ballard and Arthur Lifshin in the June 1992 issue of *ITAL* entitled "Prediction of OPAC Spelling Errors through a Keyword Inventory." Ed Goedecken's letter on their communication in the September 1993 issue of *ITAL* on his follow-up investigation of the most frequently misspelled words in NOTIS OPACs was equally intriguing. His suggestion that someone, perhaps with a large grant from CLR, should "ana-

lyze transaction logs from OPACs to discover whether these misspelled words would affect patron use of the OPAC" is a wonderful example of pure information science at its best. Here at the Molesworth Institute we have, as you might have guessed, already investigated that issue along with the related issue of how spelling mistakes by our staff affect their use of our OPAC (Online *Private Access Catalog*). By conducting a few off-the-cuff experiments, we have discovered that there is an exact correlation between the spelling mistakes in Ballard and Lifshin's [sic] list, as amplified and rearranged by Goedecken, and the spelling mistakes made by our staff

when they try to retrieve material from our OPAC and other OPACs. Using simple logic, we have come up with a fantastic solution. We have inserted a simple program in our computer that links spellings and common misspellings, in many variants, of all of the words on that list and a number of others. Thus, it does not matter if a one of those words is misspelled in an OPAC or by a person using an OPAC. All of the appropriate records are retrieved and displayed. I trust that further scientific investigation of this weighty problem will no longer be necessary.—*Collegially, Norman D. Stevens, Director, The Molesworth Institute*

---

## Index to Advertisers

3M	cover 4	LITA	2
ALA	76, 85, 96	NISO	26
Blackwell	14	PAIS	5
CD Plus Technologies	1	Todd	cover 3
Gaylord	25	WLN	50
Library Technologies	cover 2	H. W. Wilson	62

# ALA Books: Library Management Titles a Cut above the Rest



## **Basic Budgeting Practices, 2nd edition**

Richard S. Rounds

"A well thought-out presentation of a management skill every librarian must master, *Basic Budgeting* is must reading for those inexperienced in the process."

—*Public Library Quarterly* (of the first edition)

\$25.00pbk. • 1994 • approx. 180p.

ALA Order Code 0630-3-0011

## **Putting Service into Library Staff Training: A Patron-Centered Guide Library Administration and Management Association Occasional Papers**

Joanne M. Bessler

Helps library managers at all levels guide their staff in caring for library patrons and offers concrete suggestions on attracting, hiring, training, supervising, and evaluating staff who combine proficiency in their specific jobs with sensitivity to patron needs.

\$14.00pbk. • 1994 • approx. 70p.

ALA Order Code 3437-4-0011

## **Guide to Selecting and Acquiring CD-ROMS, Software, and Other Electronic Publications**

### *Acquisitions Guidelines #9*

Stephen Bosch, Patricia Promis, and Chris Sugnet

Provides the direction and suggestions for selecting and acquiring the electronic publications librarians need. From policy concerns to leasing and rental agreements, the guide demonstrates a remarkable level of detail. Selection considerations not weighed when acquiring traditional formats are included.

\$10.00pbk. • 1994 • approx. 50p.

ALA Order Code 0629-X-0011

## **Introduction to Automation for Librarians, Third Edition**

William Saffady

Describes information technology as it relates to library operations. Coverage of electronic document imaging systems, programming languages, and prewritten software packages has been greatly expanded. Discussions of CD-ROM cataloging and online catalogs have all been completely updated.

\$45.00cl. • 1994 • approx. 430p.

ALA Order Code 0628-1-0011

## **Guide to Technical Services Resources**

Peggy Johnson, editor

Describes more than 2,000 selected print, electronic, and expert sources for modern technical services practice, research, and education. Until now, no single guide covered the full range of these sources for quick access. Organizes the most important sources within such technical services categories as acquisitions, descriptive cataloging, and authority control.

\$60.00cl. • 1994 • 313p.

ALA Order Code 0624-9-0011

## **FROM THE LIBRARY AND ADMINISTRATION MANAGEMENT ASSOCIATION**

### **Library Buildings Consultant List**

Kazuko Dailey, editor

\$12.50pbk. • 95p. • 1993

ALA Order Code 0597-8-0011

### **Human Resource Management in the Small Library**

*LAMA Small Library Publications Series, #21*

Linda Sue Dobb and Patricia Dick

\$5.00pbk. • 16p. • 1993

ALA Order Code 5751-X-0011

### **Volunteers in the Small Library**

*LAMA Small Library Publications Series, #20*

Rashelle Karp

\$5.00pbk. • 16p. • 1992

ALA Order Code 5756-0-0011



ALA Books  
American Library Association  
50 East Huron Street  
Chicago, IL 60611  
To Order, call 800-545-2433  
and press 7



# TODD ENTERPRISES INC.

# NETWORK SOLUTIONS

## from 4 to 56+

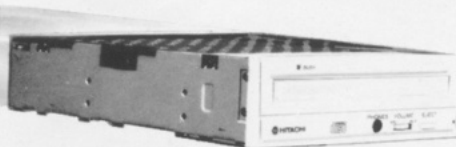
## CD-ROM Drives

### in ONE CABINET

*Todd massive data retrieval Total Access Systems feature dockable double-speed drives and security keylocks. Double speed drives have a three-year warranty. The security lock feature of Todd Total Access Systems is useful in public areas. Todd services include hardware design and development, customized systems, and network installation and management. Todd CD-ROM hardware will operate within your workstation environment with all administrating software. Daisy chain up to 256 CD-ROM drives with the Todd expansion board. Call **800 445-TODD** or (718) 343-1040 for more information.*



Showing a large 64-drive console and an 8000 series tower. Consoles available in five different sizes. Both consoles and tower with internal CPU units and dockable drives. Network towers, multidrives and consoles with your existing workstations.



Todd dockable drives slip in and out without tools or wiring. It is easy to expand your system.

since  
**1971**  
over **20** years  
OF EXCELLENCE  
**TODD**  
Enterprises

224-49 67th Avenue · Bayside, NY 11364  
718 343-1040 · FAX 718 343-9180 · BBS 516 829-0212

**800 445-TODD**

available on GSA contract

# Check out. Check in. Check in. Check out. Check out. Check in. Check in. Check out. Check out. Check in.

(Doesn't it seem like you don't have time for anything else?)

If your library is like most, the answer is probably yes. In which case you're a perfect candidate for 3M Materials Flow Management.

"Materials flow" means the actual movement of your library's "hard assets." Its books and periodicals. Audio and videocassettes. Records. CDs. Whatever you loan out and check back in.

3M Materials Flow Management streamlines this movement of materials with several innovative products.

One is the revolutionary 3M SelfCheck System, which unites your automated circulation with library security. 3M SelfCheck is the only patron self-checkout machine

that lets your patrons check out books and other items *without compromising your library's security.*

By taking care of some of your tedious and repetitious tasks, 3M Materials Flow Management boosts productivity. Makes your library more efficient. Frees staff to help patrons who really need it.

And gives you more time for the thousand and one other things that need doing in your library.

For more information on 3M Materials Flow Management or the 3M SelfCheck System, call your 3M sales consultant. Or 1-800-328-0067. In Canada, (519) 451-2500.

*The 3M SelfCheck System is a key component of 3M Materials Flow Management*

