Volume 11,
Number 1

ISSN 0730-9295

Information Technology and Libraries

March 1992

CONTENTS

ITLBDC 11(1) 1–88 (1992); Inf Technol Libr

ISSN 0730-9295

Information Technology and Libraries

Volume 11, Number 1: March 1992

CONTENTS

ITLBDC 11(1) 1–88 (1992); Inf Technol Libr

Information Technology and Libraries

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*, University of the Pacific, Stockton, CA 95211. 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*, Arizona State University, Tempe, AZ 85287. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

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 book reviews.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $20, is included in membership dues. Nonmembers may subscribe for $45 per year in the U.S.; $50 in Canada, Mexico, Spain, and other PUAS countries; $55 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 Publishing Services (Eileen Mahoney; Dianne Rooney; Amy Brown, Bruce Frausto, Josephine Gibson-Porter, Dan Lewis, and 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 the Library and Information Technology Association or the American Library Association.

Abstracted in *Computer & Information Systems*, *Computing Reviews*, *Information Science Abstracts*, *Library & Information Science Abstracts*, *Referativnyi Zhurnal*, *Nauchnaya i Tekhnicheskaya Informatsiya*, *Otdyelnyi 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 available 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.

Preliminary LC Records for Monographs in OCLC

Barbara G. Preece and Mary Anne Fox

OCLC's decision to load Library of Congress preliminary cataloging for monographs into the online union catalog resulted in the addition of a considerable number of these records to the database over a fifteen-month period before the project was suspended in March 1991. This paper presents background information on LC's decision to include these records as part of its tape-distribution service. It describes the records' impact on work flow in the cataloging department of a medium-sized research library and also reports the results of a survey that queried ARL/OCLC libraries on the use of minimal-level/preliminary cataloging records.

OCLC users were notified in February 1990 (via a logon message) that the previous month the Library of Congress (LC) had begun to include Level 5 monographic records as part of its tape-distribution service. Since it had little information as to their nature, OCLC initially processed these as "normal records" until members complained that some of them had replaced member copy. More than 900 records were affected out of the first 6,800 loaded. OCLC quickly revised its tape-load procedures so that a Level 5 record was added only if matching member copy was not in the database; if a match occurred the Library of Congress' "DLC" holding symbol was added to the existing record.

In a network newsletter issued shortly thereafter, OCLC acknowledged member complaints and warned that "no assumptions can be made about whether the [Level 5] record is correct . . . no authority work has been done, nor subject analysis provided."[1] These brief records had no call numbers, no added entries, and no notes; they occasionally lacked series statements, and some contained typographic errors (see figure 1). The descriptive cataloging elements for Level 5 monographic records are even briefer than those required by *National Bibliographic Record—Books* for standard minimal-level cataloging. They include only the following data:

1xx: first personal name on the title page
245: title and statement of responsibility in full
250: full edition statement
260: first place and publisher, one date except for multipart items
300: simplified description
4xx: series in full as it appears on the piece, whether traced or not
5xx: acquisitions data
020: first or most appropriate ISBN
010: LCCN supplied
050: IN PROCESS note supplied
Fixed fields: language, country of publication, priority[2]

Responding to those who might be concerned about the prospect of using such bare-bones records for copy cataloging, OCLC promised to monitor the situation and to consider "additional changes to rec-

Barbara G. Preece is Head, Online Catalog Maintenance, Cataloging Department, Morris Library, Southern Illinois University, Carbondale, and Mary Anne Fox is Head, Original and Special Materials Cataloging, Cataloging Department, Morris Library, Southern Illinois University, Carbondale.

SAMPLE PRELIMINARY RECORD

```
OCLC: 22489311   Rec stat: n   Entrd:  900129        Used: 901011
Type: a Bib lvl: m Govt pub:   Lang:   eng Source:   Illus:
Repr:   Enc lvl: 5 Conf pub: 0 Ctry:   idu Dat tp: s M/F/B: 00
Indx: 0 Mod rec:   Festschr: 0 Cont:
Desc: a Int lvl:   Dates: 1990,
  1 010      90-80260
  2 040      DLC | DLC
  3 020      0884300943 (pbk.)
  4 050 00   IN PROCESS (ONLINE)
  5 092       |b
  6 049      SOII
  7 100 10   Hagen, Lyman B.
  8 245 10   Dee Brown / |c Lyman B. Hagen.
  9 260 0    Boise, Idaho : |b Boise State University, |c c1990.
 10 300      52 p. ; |c 20 cm.
 11 440  0   Boise State University western writers series ; |v
             no. 95
 12 500      PRIORITY 3

This record duplicated a minimal level record (OCLC 22357599).
It was replaced by complete LC cataloging which included the
addition of 5 subject headings, two added entries, a note, and
correction to the series statement.
```

Figure 1. Sample Preliminary Record.

Figure 1. Sample Preliminary Record.

ord processing." Meanwhile, members had the option of upgrading Level 5 copy to K- or I-level standards.

LC's plan to share information about its recent acquisitions via its tape-distribution service dated back a number of years. The Technical Services Directors of Large Research Libraries (TSDLRL), which met at the ALA Annual Conference in Dallas in 1984, reacted favorably to LC's proposal to distribute in-process data.[3] This proposal came at a time when some research libraries also were beginning to consider standardized minimal-level record creation as an effective means of attacking cataloging backlogs and promoting resource sharing.[4] A 1985 report prepared for the Association of Research Libraries (ARL) Committee on Bibliographic Control explicitly recommended inclusion of LC in-process data in the national utilities. The expectation was that this information would decrease duplicate cataloging efforts by indicating the relative priority given to a work by LC. The report's authors also believed that the appearance of in-process data would allow libraries to "predict when LC copy may be forthcoming."[5]

In the late 1980s LC reevaluated its methods for assigning cataloging priority levels to incoming materials (including monographs, serials, and microforms). The revised guidelines were published in *Cataloging Service Bulletin* (*CSB*) in the 1991 winter issue. LC reemphasized an item's intrinsic research value, i.e., its content, over any former preference for language or source of material. LC also provided detailed information on the terms used to evaluate incoming materials and to place them in one of four cataloging priority levels. The decision in each case was to be based upon a careful balance of factors specific to the item, as well as LC's internal needs. *CSB* 51 also featured a brief description of LC's preliminary, or in-process, records. LC cautioned, as OCLC had, that "No assumptions can be made as to whether headings in the record reflect established forms, nor whether the record meets the National Level Bibliographic Record minimal-level cataloging specifications."[6] Neither article specified when more complete cataloging would be supplied. LC expected that "once the revised priority assignment procedures have been operational for a reasonable time, the Library will periodically provide figures on how long items

in each priority category take to go through the series of required processing steps."[7] Its plan was to catalog completely items in priority levels 1–3; those assigned to Priority 4 and titles remaining in its cataloging arrearage for three years or more would be candidates for minimal-level cataloging.

STUDY

Because previously loaded preliminary records will remain in the OCLC database for the foreseeable future, it may be useful to describe the problems experienced in dealing with these items by the cataloging staff of a medium-sized research library. The data presented here are from a project at Morris Library at Southern Illinois University, Carbondale, designed by the authors.

From July through December 1990, we reviewed all Level 5 records identified during OCLC searches of monographic titles entering the cataloging department. We recorded the OCLC record number and the item's priority level, general subject area, language, place of publication, and number of holdings. Following the test period, we searched these records again to see whether they had been upgraded, merged with a duplicate record, or replaced by LC complete cataloging. We also noted any holding changes.

Of ninety-two records encountered during the six months, thirty-nine (42 percent) were eventually upgraded by Morris Library catalogers.[8] Other OCLC members upgraded twenty-eight records (30 percent). Fourteen records (15 percent) had been replaced by full DLC copy, and duplicate record mergers had occurred four times. Seven items (8 percent) had not been upgraded when the records were researched in February 1991. The six Priority 4 records replaced by LC had received complete cataloging. Rather surprisingly, LC had replaced twice as many Priority 4 records as it had Priority 3s.

Using the before-and-after versions of the records, we compiled a profile of the group (see figure 2 and table 1). More than one-third of the records had an original priority assignment of Level 4; 10 percent had been assigned Priority 2. The "age" of the records at the time we first encountered them can be summarized as follows: Priority 2, two weeks to nine months; Priority 3, two weeks to twenty-eight months; Priority 4, two weeks to eight months. Since "date of entry" in the fixed field was sometimes as early as 1988, that element obviously referred to the date LC had created the record rather than when it entered the OCLC database.[9]

We assigned each item to one of three subject areas: social studies (sixty-eight

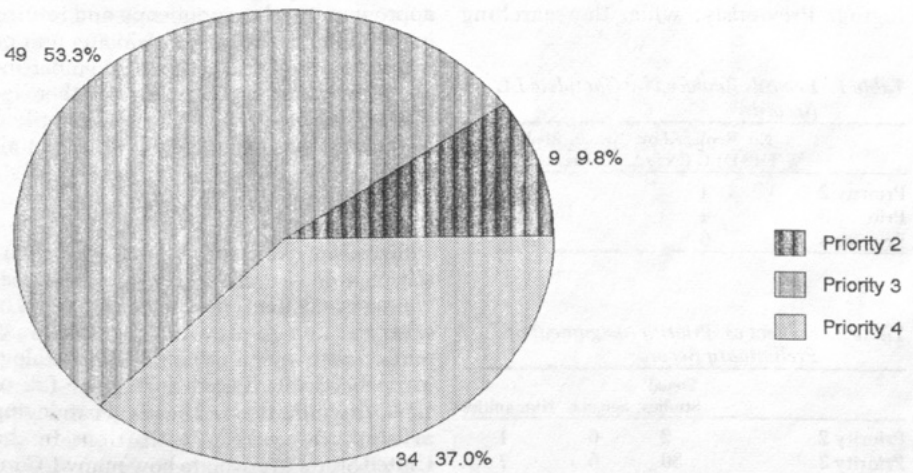

Figure 2. Preliminary Records by Priority Level.

items), science (twelve items), or humanities (twelve items). These were subsequently matched with the priority levels assigned by LC. We found that most humanities and social studies items were placed in Priorities 3 and 4, while science items were in Priorities 2 and 3 exclusively (see table 2).

Items in foreign languages (chiefly German) accounted for 38 percent of the entire sample. No Priority 2 items were in a foreign language. Out of the thirty-five foreign-language items, 86 percent were Priority 4, 14 percent were Priority 3. While only a third of the Priority 2 records were for items with a non-U.S. imprint, foreign imprints accounted for nearly half of Priority 3 items and for almost all of those in Priority 4. Foreign imprints made up 65 percent of the entire sample (see figure 3).

As noted earlier, the authors compared the number of holdings on a record at the beginning and end of the project. The average number of OCLC holdings at the point of receipt was 3.5; following the test period the average had risen to 20.6.

DISCUSSION

The problems our catalogers had with these records may help to illustrate why this experiment came to an abrupt end. For instance, 20 percent of our initial searches yielded Level 5 records that duplicated member input full (or minimal-level) cataloging. Previously, when the searching staff found duplicate records they were instructed to select DLC copy. Now it was necessary to have all records printed if one of them displayed encoding Level 5. This additional searching was the first delay in cataloging of these items.

Table 1. Records Replaced by Complete LC Records

	No. Replaced by Full DLC Record	% Replaced at Particular Level
Priority 2	4	44
Priority 3	4	8
Priority 4	6	17

Table 2. Subject vs. Priority Assignment of Preliminary Records

	Social Studies	Science	Humanities
Priority 2	2	6	1
Priority 3	36	6	7
Priority 4	30	0	4
Percentage of Total	74	13	13

The library received fifteen to twenty items with preliminary copy each month. It did not seem practical to defer cataloging until LC supplied full records, nor did we wish to impede user access for an indefinite period. Since the cataloging department had already developed procedures for upgrading minimal-level records, Level 5 records were integrated into this work flow.[10] This meant that after searching, Level 5 items were immediately routed to the original cataloging unit.

Catalogers discovered that items with Level 5 records required almost as much time to catalog as those without copy. In addition to authority work performed for name, subject, or series entries added to the record, catalogers had to verify any existing main entries, because these headings frequently were not in *AACR2* form. They also assigned a Dewey call number, added notes, and revised the fixed field as appropriate. The LC priority designation and the message "in process (online)" were deleted. All this took considerable time for an item with ostensible cataloging copy. Additional effort was required if duplicate records existed. Catalogers had the option of either approving member input copy and returning the item to the copy cataloging unit or upgrading the DLC record if member input copy lacked access points or other significant information. OCLC was notified of the existence of duplicate records in all cases.

ARL SURVEY

Since Morris Library's catalogers generally regarded Level 5 records as a nuisance, we were curious to know what catalogers in other research libraries thought of them. As part of a survey on minimal-level cataloging policies distributed during the fall of 1990, the authors asked heads of cataloging at ninety-nine ARL institutions in the United States to estimate how many LC in-process records were encountered monthly, how they were edited, and whether they

were viewed as a problem (see appendix A). Since this study focuses on Level 5 records used for cataloging in OCLC, we excluded twenty-two institutions known to be Research Libraries Group (RLG) members. This left a potential pool of seventy-seven ARL/OCLC libraries. Fifty-seven (74 percent) of these responded to our survey. One of the fifty-seven responses was incomplete and was excluded.

The responses showed that 60 percent of the OCLC libraries had formal policies on the creation and use of minimal-level cataloging; 50 percent used some form of minimal-level cataloging themselves. Eighty-four percent indicated that their libraries upgraded minimal or preliminary records in OCLC; of this subgroup 80 percent reviewed the entire record beforehand, rather than selectively checking and modifying fields. Level 5s were upgraded by 79 percent of the respondents. Notes, subject headings, and added entries were the fields added most often to LC preliminary records. The "in process (online)" message was usually not deleted by the editing library. Sixteen libraries (29 percent) reported seeing more than fifty Level 5 records a month, and several said they encountered hundreds per month.

Various methods of editing and/or upgrading preliminary or minimal-level records were cited. A few libraries edited the record without upgrading; twenty-five (45 percent) upgraded in compliance with OCLC Bibliographic Input Standards, while fourteen (25 percent) combined or varied these methods, depending on the nature of the material being cataloged. Professional catalogers processed these materials 50 percent of the time. Two percent of the libraries used paraprofessionals exclusively, while 48 percent used a combination of both professionals and paraprofessionals, depending on the type of material.

Respondents identified many of the same drawbacks in these records that our catalogers had discovered: they created more work, especially for higher-level staff; they resulted in the creation of backlogs and disruption of the work flow; DLC records could no longer be considered authoritative as a matter of course. Some libraries reported that they did not have the time to

Figure 3. Language and Publication Data of Preliminary Records.

upgrade all the Level 5s they encountered. Others held items for as long as a year in anticipation of a complete record. Opinions on Level 5 records were very negative overall. Only three respondents commented favorably on them, expressing the viewpoint that "something is better than nothing."

CONCLUSIONS

Whatever their contribution to collection development and resource sharing might have been, the value of LC in-process records as sources of cataloging copy appears to have been negligible, chiefly because of their duplicative nature and the special routines needed to handle them. Their inclusion in a national database subverted the long-held notion that DLC records could be used with little editing.

Several observations may be made concerning the sample records that we studied. We found a preponderance of foreign language and foreign imprints had been assigned to the lowest cataloging priority, despite LC guidelines that no longer discriminate in favor of English language or domestic publications. The substantial growth in holdings per record during the test period indicates that research libraries are likely to acquire—and require cataloging copy for—many items considered by LC to have low research value and need. If a particular library cannot afford the time or staff to upgrade or fully edit Level 5 records, the ultimate result will be less access in an online catalog.

In April 1991 OCLC announced that it had stopped loading LC in-process monographic records because of widespread complaints about their quality and because of continuing technical difficulties in matching them with existing member copy. At that point approximately 171,000 Level 5 monographic records remained in the database.[11] Since many had multiple holdings attached, they could not be deleted. OCLC expected their number to diminish gradually as they were upgraded by member libraries, merged, or replaced with more complete LC copy. An experiment designed to foster resource sharing had foundered because of its incompatibility with the needs of network members for a standardized and reliable cataloging source.

REFERENCES AND NOTES

1. "Encoding Level 5 Records in the OLUC," *Bits and Pieces*, no.116:2–3 (Mar. 1990).
2. "Simplified In-process Record Specifications," addendum to Sept. 13, 1990, memorandum from Duane E. Webster, Association of Research Libraries, to directors of ARL Libraries.
3. Colleen Bednar, Library Exchange, *RTSD Newsletter* 9, no.6:75–76 (1984).
4. Michael Gorman, "Report on the Technical Services Directors of Large Research Libraries Survey on Minimal Level Cataloging," *Information Technology and Libraries* 3, no.4:382–84 (Dec. 1984); Nancy E. Douglas and Shirley Leung, "Use of the Full MARC Record: "Myth and Reality," in *Academic Libraries: Myths and Realities*, ed. Suzanne C. Dodson and Gary L. Menges (Chicago: Association of College and Research Libraries, 1984), p.177–82.
5. Carol A. Mandel and Susan F. Rhee, "Shared Cataloging: Some Remaining Issues," *Cataloging & Classification Quarterly* 7, no.2:33,37 (Winter 1986).
6. *Cataloging Service Bulletin* 51:50 (Winter 1991).
7. Ibid., p.4.
8. This number includes two "reupgraded" records in which it was apparent that the in-process record had bumped member copy. It was necessary, among other things, to reverify and correct the main entries for these items.
9. This is with the exception of the items referred to in note 8.
10. Mary Anne Fox and Barbara G. Preece, "Upgrading Minimal Level Monographic Records: A Study and Conclusions," *Technical Services Quarterly* 8, no.4:25–35 (1991).
11. Personal communication with OCLC Office of Quality Control; ILLINET/OCLC Service. *Information Bulletin*, no.230:2 (Apr. 15, 1991). ■■

APPENDIX A. SURVEY ON MINIMAL/PRELIMINARY CATALOGING AT ARL/OCLC LIBRARIES

Please circle your response to the following questions.

Part I. General Information

1. Does your library have an online catalog? (If no, please stop here and return the survey.) Y N
   1a. Does your local system provide keyword title access? Y N
2. Does your library have a formal policy on the creation of minimal-level cataloging records for your online catalog? Y N
3. Does your library have a formal policy on the use of minimal-level cataloging records for your online catalog? Y N
4. Does your library contribute minimal-level cataloging records to OCLC? Y N
   4a. If yes, check the type of material that receives minimal-level cataloging (check all that apply). If no, go to question 5.
   Theses/dissertations _____ Maps _____ Kits _____
   Archives/manuscripts _____ Microforms _____
   Sound recordings _____ Computer files _____
   Other (please specify) _____

Part II. Upgrading

5. Do you upgrade minimal- or preliminary-level records in OCLC? (If no, go to question 8.) Y N
6. Do you upgrade: Level K records? Y N
   Level M records? Y N
   Level 5 records? Y N
   Level 7 records? Y N
7. When you upgrade a minimal- or preliminary-level record, do you review the entire record or do you check only specific fields?
   Entire record _____ Notes _____ Description _____
   Access points other than subject headings _____
   Subject headings _____ Fixed fields _____

Part III. Preliminary Records

8. Estimate how many Level 5 records you encounter per month:
   0–15 _____ 16–30 _____ 31–50 _____
   more than 50 _____
9. If you upgrade or edit Level 5 records, which of the following fields do you add? (Check all that apply.)
   090 _____ 092 _____ 500 _____ 504 _____
   6xx _____ 7xx _____
10. Do you delete any fields on Level 5 records? Y N
11. Do you need more information concerning LC's use of Level 5 cataloging? Y N

Part IV. Editing

12. When copy cataloging with minimal-level cataloging do you:
    Fully edit the OCLC record without upgrading? Y N
    Upgrade the record in accordance with BIS? Y N
    Use the record as is, adding call number and holdings? Y N
13. Who edits the minimal-level record in preparation for the online upgrade?
    Paraprofessional Y N
    Professional Y N
14. Do you view the increase in the number of minimal-level cataloging records in bibliographic utilities, such as OCLC, as a problem? Y N
    If yes, please comment.

15. Other comments.

Mainstreaming Data: Challenges to Libraries

Diane Geraci and Linda Langschied

Libraries are increasingly aware that their role includes providing access to data in electronic form. Determining the level of service and acquiring the skills needed to provide effective access are key to integrating data services with success in any organization, as is having the administrative support to do so. The proliferation of electronic media, formats, hardware, and software requires new knowledge bases. Libraries that do not take a leadership role will forfeit a pivotal position in assisting patrons with accessing the wealth of information available in electronic form.

For the past several years, U.S. academic libraries have been anticipating the inevitable—the receipt of large amounts of 1990 census data in electronic form. Statements of why libraries should be involved in providing access to computer-readable census data are not new, as this quote from Rowe and Ryan (1974) illustrates:

> Why not just store the tapes at the computer and let the computer people handle them? By doing this, the library would be abdicating its role as an information center. It would deny users the opportunity of locating information at the place we have trained them to look for it, the library.[1]

What *is* new for the 1990s is the complication of a greater variety of electronic format, software, hardware, and network decisions to consider.

The growth of involvement by academic libraries in the realm of computer-readable data, while slow in coming, has been incremental. Many of the libraries that formerly eschewed responsibility for providing services to computer-readable data are now facing the issues surrounding data files. The infusion of CD-ROMs containing numeric data in libraries as part of the U.S. Federal Depository Library Program has assisted in placing a sense of urgency among those receiving them. The time has come for more libraries to consider mainstreaming data services, while keeping in mind that there is a wealth of experience from which to draw in the data library and archive community that is already well established across the United States and Canada.[2]

WHAT CONSTITUTES DATA, AND HOW ARE THEY USED?

"Words, Pictures, Numbers, and Sounds: Priorities for the 1990s" was the theme of the International Association of Social Science Information Services and Technology's (IASSIST) 1990 annual meeting and illustrates the breadth of data usage for research and teaching.[3] Data in electronic form can include public opinion surveys, hospital-admission records, digital cartographic data, literary works, digital storage of sound bites, video footage, photo-

Diane Geraci is Social Science and Data Librarian, Bartle Library, State University of New York at Binghamton, and Linda Langschied is Information Services Librarian, Alexander Library, Rutgers University, New Brunswick, New Jersey.

graphs, and much more. Computer-readable data are used routinely in the humanities, social sciences, and sciences. Libraries cannot afford to neglect them, but rather need to understand how data are created and how they are used in order to respond to new service demands.

Social Data

Computer-readable social data are a proliferating body of information and research sources, derived from the surveys, censuses, and administrative records of a multitude of commercial research groups, government agencies, academic institutions, and private research agencies. The amount of numeric data produced and available for secondary data analysis has dramatically increased and is distributed in a variety of computer-readable formats. Quantitative analysis of data is one of the essential methodological approaches used in a variety of social sciences and related disciplines, including anthropology, business, economics, education, geography, health-care fields, history, political science, psychology, and sociology.

The scope of data available for secondary analysis has also grown tremendously. Common categories of data include economic, political, and social attitudes and behavior patterns; social indicators and quality of life; business and commerce; population and housing; education; employment; aging and life cycle; crime and criminal justice; and health care and health facilities. Practically any social phenomenon that can be measured or counted is fair game for social-research use.

The use of existing data files to raise and answer questions that did not occur to the original researcher is both legitimate and cost-effective. The wealth of data gathered in one survey is rarely thoroughly explored by the principal investigator. Data gathered in another survey may be merged with administrative records to create a new compilation of data with more exciting possibilities for analysis. More students, both undergraduate and graduate, are being trained in survey research methodology, quantitative analysis, and computer literacy. The ability to use a well-documented collection of data to test new ideas, to explore policy implications, or simply to extract information not available elsewhere is no longer restricted to a small group of researchers.

Data in computer-readable form are an increasingly prevalent phenomenon for many categories of information collected by the U.S. federal government as well as by individual researchers and private agencies. Many requests for information about government statistics lead to an electronic source. While all such requests may not require access to an entire data set or even a subset for secondary analysis, a particular question may require the extraction of several data elements for an answer. Social-data collections are sources for research use, e.g., secondary analysis, and for reference use, e.g., informational sources.

There are already several well-established data archives and distributors of social data. Among the most well-known archives are the Inter-University Consortium for Political and Social Research (ICPSR); the Roper Center for Public Opinion Research; and the Center for Electronic Records, National Archives and Records Administration. There is also a host of smaller data archives at universities in North America that house more specialized collections. Among them are the Henry A. Murray Research Center of Radcliffe College, Louis Harris Data Center at the Institute for Research in Social Science at the University of North Carolina at Chapel Hill, the National Data Archive on Child Abuse and Neglect at the Cornell Institute for Social and Economic Research at Cornell University, and the Research Resource Division for Refugees in the Centre for Immigration and Ethno-Cultural Studies at Carleton University.

Science Data

Computer-readable data in the scientific community encompass a wide range of resources including chemical and physical reactions and properties data, engineering and materials data, geoscience and astronomical data, and life sciences and biosciences data. The quantity of scientific data has grown significantly within the last twenty years with the assistance of increasingly sophisticated instruments for

measurement and computers for storage and analysis. These advancements allow more users of science data to rely on data collected by others rather than to rely solely on the results of their own experiments and observations.

The physical and chemical sciences generate data that are derived from well-defined repeatable experiments. For instance, high-energy accelerators produce millions of particle-collision events that serve as primary research information for physicists. The geosciences and astronomy tend to produce observational data, measuring unique events that are not easily repeatable.[4] Observational data include measurements of seismic activity, quantities of mineral resources, the positions and movements of astronomical objects, measurements of temperature fluctuations or groundwater levels, images of geological features, and a host of other equally disparate observable items. Still other disciplines, like the life sciences and biological sciences, rely on descriptive data that may be nonnumeric, such as descriptions of the chemical and physical makeup of DNA and RNA sequences.[5]

Some scientific data products are based on or grew out of print sources. The computer-readable versions not only offer rapid access to data, but enhance possibilities for searching and manipulating. For example, some electronic products offer calculational capabilities that allow for property estimation, extrapolation, and curve drawing.[6] Others allow for correlating and comparing observations in ways that simply are not feasible with print sources.

Data sources in the sciences include government agencies, universities, research institutes, as well as commercial vendors. Government agencies are particularly rich resources for archiving and distributing science data. The National Oceanographic and Atmospheric Administration (NOAA), the National Aeronautics and Space Administration (NASA), and the U.S. Geological Survey (USGS) are three such major purveyors. There are also many sources of commercially available scientific data. Examples include Satler Spectra, a computer file of graphic representations of electromagnetic radiation absorbed or emitted by molecular or atomic fingerprints, and Powder Diffraction Data, a collection of X ray or electron-diffraction crystal-structure data of chemical compositions.[7] Overall, the rapidity with which science computer-readable data files are produced has, so far, defied attempts to manage them effectively and to make them more accessible to the larger scientific community.

Humanities Data

Are data restricted to numbers only? This is a common conception, but computing humanists have helped to extend the definition of *data*. For example, the computer might allow a researcher to reduce the melodies of Mozart to bits and bytes in order to compare musical sequences,[8] to analyze large lexicons for natural language processing,[9] or to analyze the lexical structure of Shakespeare's sonnets.[10]

The most widespread use of data analysis in humanities research lies in textual analysis. Words as data, a notion that arose with concordance building of the works of St. Thomas Aquinas in the late 1940s, has become an increasingly accepted concept by linguists and literary scholars. Computational humanists have faced a major obstacle that their counterparts in the social sciences encounter less often: a lack of existing data files to analyze. With no readily available corpus of machine-readable texts, humanists often must create their own data sets for analysis—in itself no small task. In his 1984 stylo-statistical study of Benjamin Constant's nineteenth-century French romantic novel, *Adolphe*, Professor Robert Allen reveals that laying the groundwork for his study took approximately seventeen years: twelve years to write the programs to extract data from texts and up to five years to encode a single text.[11]

Several academic initiatives have been undertaken to make humanities text files accessible. Established in 1981, the American and French Research on the Treasury of the French Language (ARTFL) Project is a cooperative effort of the Centre National de la Recherche Scientifique and the University of Chicago. The ARTFL database, a collection of nearly two thousand French-language texts, is available to sub-

scribers through remote online access. Another online system, the Dartmouth Dante Project, provides a database of more than six hundred years of commentary on Dante's *Divina Commedia*. The Thesaurus Linguae Graecae Project, based at the University of California, Irvine, reproduced the entire Greek corpus from Homer through the sixth century A.D. utilizing CD-ROM technology.

The Oxford Text Archive at Oxford University was established in 1976 to offer scholars storage and maintenance of electronic versions of literary and linguistic texts and to manage the distribution of those texts to the scholarly community. In 1991 Princeton and Rutgers universities created a national center for machine-readable texts in the humanities to serve as a central depository to aid researchers in locating and procuring existing texts and reduce needless duplication of data entry and encoding. Housed at Rutgers, the Center for Electronic Texts in the Humanities (CETH) will inventory and catalog existing machine-readable texts, provide the inventory in an online environment, preserve and provide access to the data files themselves, and compile texts on demand.[12]

Another current initiative that promises to enhance access to machine-readable texts for humanist scholars further is the Text Encoding Initiative (TEI), jointly sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing. The TEI promotes uniform minimal encoding standards for text files.[13] Encoding standards for texts will greatly facilitate secondary analysis and wider use of text files for humanists.

WHY MAINSTREAM DATA?

Whether planning for the information and research needs of students, social scientists, scientists, or humanist scholars, libraries need to begin to consider the status of computer files within their institutions. As Johnson notes:

> . . . computer files have grown in their importance to research while research libraries and librarians have largely continued to ignore them. As a consequence, many researchers have become accustomed to circumventing the library and have devised their own methods for gaining access to such files through academic departments and direct contact with producers and data archives.[14]

Libraries must acknowledge and accommodate this increasingly important body of electronic resources in order to continue to be relevant providers of information in the future.

Many library users require computer-readable formats to accomplish what they need to do. The enhanced capabilities for manipulation offered by computers and various software packages provide an incentive that is difficult to ignore. Actual mainstreaming, or integration of computer-readable data into our tool kit of information sources, may indeed pose a challenge. However, we need to evaluate whether we can afford to disregard many of the resources that are currently readily available, such as those listed below.

- Familiar library reference tools—for example, ERIC—cite computer-readable files. Should we end the quest there, or should we be able to provide further access?
- *The World Factbook* and *County and City Data Book* are available on CD-ROM, diskettes, magnetic tape, and in print. Should we be able to direct patrons to the appropriate source for their needs, as well as assist them in extracting information?
- Print codebooks, i.e., technical documentation that accompanies computer data files, are information sources in and of themselves. For example, a codebook may include frequencies of responses to survey questions and can serve as a model for constructing survey questionnaires. How can we integrate these valuable sources into a typical reference service?
- The U.S. federal government has moved to distribution of some government data in electronic form only, for example the *U.S. Imports of Merchandise* and *U.S. Exports of Merchandise*, and large portions of the 1990 decennial census. How do we integrate access to these resources into our current services?
- Many humanities texts are already available in electronic form, for example, the King James version of the *Bible*. Should

we at least be able to identify electronic texts, if not acquire them, and direct users in their use?

- Many science print-based resources have moved to electronic format—for example, the Environmental Protection Agency's *Toxic Release Inventory*, formerly available on microfiche. Can we direct researchers to make use of the expanded capabilities of these electronic products?

CHALLENGES LIBRARIES FACE

The sheer quantity of computer-readable data in the social sciences, sciences, and humanities that is being produced, distributed, and archived for use by others has grown exponentially in the last decade. How libraries choose to incorporate data services into their daily routines and coordinate these services with others on campus are complex issues. Many of the challenges are still before us. We must clearly define our role and the services we intend to provide and acquire the skills we need to integrate data services effectively into our organizations.

Defining Ourselves and Our Services

For the uninitiated library, the question "Where do we even start?" is a real and significant one. There is no one model for data services that will work equally well for all organizations, and most organizations will rely on more than one unit to provide a full range of services.[15] Libraries must evaluate their own services, organization, and capabilities, as well as those of their parent organizations, to begin to address the question.

Although local conditions and historical developments may have determined and shaped the current role of the library, these need not dictate future roles. Libraries can build new linkages to existing services on campus and determine what new roles the library can assume to improve or establish access to data sources. Bringing data into the mainstream of information services on campus should be a task central to the library's mission, although it may ultimately share the actual responsibility for providing data services with other units on campus. Because the library serves all constituencies on campus and because its mission explicitly includes providing access to information, the library is a logical and appropriate unit to coordinate service for access to data resources.

Since no one organizational paradigm will suit all situations, service levels must be determined within the institutional framework. Jim Jacobs, data services librarian at the University of California at San Diego, has proposed a range of services that an academic library might offer, depending on local circumstances. He makes three noteworthy assertions: (1) most libraries can provide some kind of data service without providing complete service for all conceivable combinations of users and electronic products, (2) libraries should attempt to deal with these data sources in some way, and (3) it may not be necessary or prudent for the library to provide full service for data on its own. Jacobs provides four ranges of services to assist those interested in determining the appropriate level of service for their institution. They are (1) levels of data services, (2) levels of library data services, (3) levels of reference service for data files, and (4) levels of computing services.[16] While the levels are not meant to be prescriptive, they illuminate many of the points that need to be considered when libraries are assessing their own role in providing data services.

Libraries should establish a written service plan that outlines the services to be provided and indicates which units within the library and the larger campus community will provide them. Units that might participate in the provision of services and contribute to such a service plan include the computer center, a data archive, a statistical consulting service, a research institute, and various academic departments. Administrative delineation of which units are responsible for which services removes ambiguity for both staff and users and fosters cooperation and communication among campus units that may not report through the same administrative channels. In addition, a written service plan helps avoid problems that might otherwise arise when individuals who informally agree to be responsible for services later move to new po-

sitions. Finally, a service plan alleviates the duplication of services among different units that campus budgets simply cannot support.

In addition to outlining the library's role within the campus community, the internal operations of the library must be evaluated with regard to data-service provision. A careful assessment of current library services is crucial as data services will cut across traditional library departments such as reference, collection development, cataloging, acquisitions, government documents, and online services. These services also may span branch library units such as undergraduate and graduate libraries, humanities and social science libraries, and even medical and science libraries. Bringing data service into the mainstream necessarily entails the involvement of many library personnel, and their efforts should be well planned and implemented in a concerted manner.

Expanding Our Knowledge Bases

Information skills, computing skills, and knowledge of research methodologies come together to provide data services for humanities, social science, and science computer files. Many library and information skills are transferable to servicing computer-readable data. These include cataloging, reference interviewing, online searching, acquisitions, personnel management, fiscal management, and collection development and management. Subject specialists may bring their knowledge of the research methodologies used in their fields to bear in collecting data and in assisting researchers in utilizing them. Other skills, particularly computing skills, may need to be developed. Gaining new skills through self-training, continuing-education courses, locally offered computing classes or discipline-specific methodology courses, and membership in appropriate professional organizations will be necessary to keep abreast of these new technologies and new service demands.[17]

Notions that computing skills are the domain of the library systems staff or computing center staff impede the development of library data services as well as many of the other services libraries provide. The division between "library skills" and "computing skills" has become less distinct.

The complexities of technical issues involved in servicing data may, however, pose the greatest difficulty in fully integrating data services into the library environment. It is no coincidence that in the previous three decades, most data-service operations for social scientists were established outside of libraries.[18] The medium itself has been a stumbling block for many libraries because "the computer file, a dynamic medium . . . does not mesh neatly with existing library collections, structure, routine, and services."[19]

Unfortunately, neither users nor data-support staff benefit from isolating computer-readable data from similar resources that reside in the library.

In order to offer data services, librarians will need to be familiar with the attendant issues related to hardware, operating systems, applications software, networking, data format, and the variety of computing environments that exist on campus—whether mainframe, microcomputer, or workstation based. While technical configurations will vary from place to place, it is likely that librarians providing service to data will need to be conversant with a broad variety of technologies as equipment, storage media, and software proliferate. Even if librarians do not provide direct technical assistance with accessing data, they will need to be aware of different computing environments and the technical requirements of data products in order to accommodate user needs and to carry out reference, referral, and collections services.

Recently, we find data produced in various formats and distributed in a wide variety of media—e.g., paper, microfiche, diskette, CD-ROM, magnetic tape or cartridge, and online via telecommunication links and via newer high-speed networks, such as the Internet. Understanding the interrelationship among varying electronic-distribution media; technical modes in which data are written, such as ASCII and EBCDIC; and technical formats such as card image, logical record length, and

software-specific formats like dBase or SAS, is necessary for providing responsible services to data.

Perhaps the most significant new impetus for involving libraries in data services is the U.S. Government Printing Office's distribution of numeric data on CD-ROM through the Depository Library Program. CD-ROM–based data hold the promise to fill a gap between static print sources and the more difficult-to-use magnetic tape and may well stimulate a new data-user clientele. Data available in a nonmainframe environment, bundled with user interface software, may open avenues of access to users unable to master the complexities associated with mainframe technology on campus: the need for a computer account, the necessity of learning job control language (JCL), and a statistical software package like SAS or SPSS-X. These tasks are especially daunting for the occasional or one-time user. However, the medium has limitations. Extracting data from CD-ROMs can be very slow. Many microcomputers cannot easily handle the large numeric data sets that are issued on CD-ROM. Librarians must be able to assess the user's needs, with an understanding of the issues of data, hardware, and software, in order to make referral not only to the correct data, but to data in the appropriate medium and format.

For librarians to neglect the greatly expanded retrieval capabilities of electronic versions of data will surely do a great disservice to many library patrons. Would we, for example, direct a researcher wanting to study word occurrence in Shakespeare's writings to the print versions of his works and suggest taking a manual count? Are we savvy enough in data issues even to conduct an appropriate reference interview, recognizing a couched request for numeric data and referring the patron to a computer file when appropriate, rather than a print statistical source? Are we prepared to direct a patron looking for changing land-use patterns in Southern California to satellite-imagery data, even though we may not house the collections locally?

The varying sources and formats of data create complexities for collection development and reference services and, indeed, for all related library services. Library administrative support and understanding are the crucial links in establishing the appropriate structure for data service and in determining the level of service on which an appropriate training schedule should be based. In an overview of training issues for computer files, Chiang outlines the specific skills and expertise needed to perform various tasks and provide specialized services. She concludes by indicating that libraries must train librarians to offer access to computer files, "Indeed, they will have to."[20]

CONCLUSION

In 1977, Ferguson noted a number of unresolved problems concerning the involvement of libraries with data files:

> Will there be a reference librarian who is a specialist in data files . . . ? If there is a data services librarian, what percentage of time can reasonably be allowed to data file activities and for other reference activities such as collection development and general reference service? . . . How can some of the knowledge of the data services librarian be passed on to other librarians so that the service frontier is broader and the service depends less on a single person? . . . If a librarian works extensively with a computer center staff, does this create an image of a "marginal" person who is neither a programmer accepted by computing specialists or a librarian accepted by fellow librarians?[21]

Nearly a decade later, Ferguson's questions remain pertinent. How libraries will define their roles as providers of data services may still, for many, be in question; however, that libraries must provide some level of data service should no longer be in dispute.

Libraries are, in fact, in a pivotal position to provide or coordinate the wide variety of information and research services data users require. The resources of the academic library itself—a team of experienced reference personnel, reference and government documents collections, subscriptions to various online services, collection development and management professionals—all compose an excellent infrastructure for the provision of information services. The decision to extend existing services to data in computer-readable formats is not without challenges, but neither is it without precedent. The library is typically better

equipped to provide user services than many other units on campus. Many traditional librarian skills are transferable ones in the data arena. Most important, libraries are service oriented, dedicated to providing access to and organizing information in all of its myriad forms. Librarians need to define clearly the appropriate level of service and acquire the necessary skills to bring data services into the mainstream.

REFERENCES AND NOTES

1. Judith S. Rowe and Mary Ryan, "Library Service from Numerical Data Bases: The 1970 Census as a Paradigm," *College & Research Libraries* 35, no.1:7 (Jan. 1974).
2. Some examples of data-services operations and approaches to services are outlined in the following publications: JoAnn Dionne, "Why Librarians Need to Know About Numeric Databases," in *Numeric Databases*, ed. Ching-chih Chen and Peter Hernon (Norwood, N.J.: Ablex, 1984), p.237–46; Ann Gerken, "What Is a Data Archive and What Should the Information Specialist Know About Managing Locally Maintained Numeric Data Files?" *Database* 11, no.4:60–65 (Aug. 1988); Ann S. Gray and Sue A. Dodd, "The Roles of Libraries and Information Centers in Providing Access to Numeric Databases," in *Numeric Database*, ed. Ching-chih Chen and Peter Hernon (Norwood, N.J.: Ablex, 1984), p. 247–62; Ray Jones, "The Data Library in the University of Florida Libraries," *Library Trends* 30, no.3:383–86 (Winter 1982); Laine G. M. Ruus, "The University of British Columbia Data Library: An Overview," *Library Trends* 30, no.3:397–406 (Winter 1982); Bliss Beckman Siman, "Machine-Readable Data Files in the Library," *Technicalities* 6, no.6:9–10, 15 (June 1986); Elizabeth Stephenson, "Data Archivists: The Intermediaries the Census Bureau Forgot—A Review Essay of 'The Role of Intermediaries in the Interpretation and Dissemination of Census Data Now and in the Future,' " *Government Publications Review* 17, no.5:441–47 (Sept./Oct. 1990).
3. Diane Geraci, "The International Association for Social Science Information Services and Technology," *College & Research Libraries News* 51, no.9:869–71 (Oct. 1990).
4. Gesina C. Carter, "Numerical Databases for Science and Technology," in *Numeric Databases*, ed. Ching-chih Chen and Peter Hernon (Norwood, N.J.: Ablex, 1984), p.15–42.
5. Constance C. Gould and Karla Pierce, *Information Needs in the Sciences: An Assessment* (Mountain View, Calif.: Research Libraries Group, 1991).
6. Linda Langschied and others, "Machine-Readable Data Files at Rutgers: A Preliminary Report of the Task Force on Numeric Data in Machine-Readable Form," Rutgers University Libraries, 1991.
7. Ibid., p.9.
8. Marcel Mongeau and David Sankoff, "Comparison of Musical Sequences," *Computers and the Humanities* 22, no.3:161–75 (June 1990).
9. Roy Byrd and others, "Tools and Methods for Computational Lexicology," *Computational Linguistics* 13, no.3–4:219–40 (July–Dec. 1987).
10. Dean Keith Simonton, "Lexical Choices and Aesthetic Success: A Computer Content Analysis of 154 Shakespeare Sonnets," *Computers and the Humanities* 24, no.4:251–64 (Aug. 1990).
11. Robert Allen, *A Stylo-Statistical Study of "Adolphe"* (Geneva, Switzerland: Campion, 1984), p.3.
12. Marianne Gaunt, "National Center for Machine-Readable Texts in the Humanities," Posted to Public-Access Computer Systems Forum <PACS-L@UHUPVM1> (BITNET Discussion Group), July 19, 1990.
13. "Full Text Encoding Guidelines Now Available," *REACH (Research & Educational Applications of Computers in the Humanities): Newsletter of the Humanities Computing Facility of the University of California at Santa Barbara* (Sept./Oct. 1990) p.4.
14. Margaret Johnson, "Adding Computer Files to the Research Library: Issues in Collection Management and Development," in *Computer Files and the Research Library*, ed. Constance Gould (Mountain View, Calif.: Research Libraries Group, 1990), p.3.
15. Margaret O'Neill Adams, "Models of Data Library Development and Information Systems: An Overview," *Library Trends* 30, no.3:337–81 (Winter 1982).
16. Jim Jacobs, "Providing Data Services for Machine Readable Information in an Academic Library—Some Levels of Service," *Public Access Computer Systems Review* 2, no.1:144–60 (1991) [electronic journal].
17. Training for data professionals is outlined in the following articles: Laine G. M. Ruus, "Training of Data Services Professionals: Past, Present, and Future," *Library Trends*

30, no.3:455–65 (Winter 1982); Judith S. Rowe and Carolyn L. Geda, "Training the Professional Data Librarian," *Drexel Library Quartery* 13, no.1:100–8 (Jan. 1977); Edward Mignon, "Numeric Data Bases in the Professional Librarianship Curriculum," *Behavioral and Social Sciences Librarian* 1, no.3:181–87 (Spring 1980).

18. Kathleen M. Heim, "Social Science Information Needs for Numeric Data: The Evolution of the International Data Archive Structure," *Collection Management* 9, no.1:1–53 (Spring 1987).
19. Johnson, "Adding Computer Files to the Research Library," p.3.
20. Katherine Chiang, "Computer Files in the Library: Training Issues," in Research Libraries Group, *Computer Files and the Research Library*, ed. Constance Gould (Mountain View, Calif.: Research Libraries Group, Inc., 1990), p.35.
21. Douglas Ferguson, "Social Science Data Files, the Research Library and the Computing Center," *Drexel Library Quarterly* 13, no. 1:79 (Jan. 1977). ■■

Cleanup and Deduplication of an International Bibliographic Database

Stephen R. Toney

A two-year project to improve the quality of a major bibliographic database resulted in corrections to 2.1 million data fields and removal of 8.2 percent of the records. Queries now retrieve 20 percent more hits than the same queries did before the project began, despite removal of duplicate records. The literature of duplicate removal is discussed, with emphasis on the trade-offs between human and computer methods.

INTRODUCTION

Description of Database

The Conservation Information Network (Network) was created by an international collaborative effort managed by the Getty Conservation Institute, an entity of the J. Paul Getty Trust. The Network consists of an electronic messaging service and access to the following three databases:

- The bibliographic database (BCIN), consisting of references to international conservation literature; all records contain abstracts
- The materials database, containing records on products relevant to conservation practice
- The product/supplier directory, containing names and addresses of manufacturers and suppliers of materials used in conservation

The institutions contributing data records are the following:

- Canadian Conservation Institute (CCI)
- Smithsonian Institution's Conservation Analytical Laboratory (CAL)
- Getty Conservation Institute (GCI)
- International Centre for the Study of the Preservation and the Restoration of Cultural Property (ICCROM)
- International Council on Monuments and Sites (ICOMOS)
- International Council of Museums (ICOM)

Approximately four hundred institutions around the world use the Network regularly to improve their conservation of cultural property, especially art, archaeological sites, buildings, and museum collections.

The Network is resident on a Control Data Corporation (CDC) mainframe managed by the Canadian Heritage Information Network (CHIN), a government installation in Ottawa.

BCIN contained about 140,000 bibliographic records when this project began. Because BCIN was initially formed through the machine conversion of diverse files from the participants, numerous anomalies, errors, and duplicate records resulted. Furthermore, differences in cataloging standards between countries and over time led to variations in style (capitalization, punctuation, etc.). These factors contributed to the need for cleanup and deduplication.

BCIN contains all of the normal data

Stephen R. Toney is President, Systems Planning, Las Vegas, Nevada.

used for identifying and describing bibliographic records, except that LCCNs are absent, and ISBNs, ISSNs, and CODENs are rare. BCIN is not in a MARC format but is stored using Information Dimension's BASIS database management system. Maximum possible record length is 15,000 bytes, although the longest record is 5,254. The shortest is 82 bytes; the mean length is 973 bytes.

Purpose of Project

The purpose of the cleanup and deduplication project was threefold:

1. To locate and correct data errors automatically, using computer programs
2. To flag records with data errors that the programs could detect but could not correct
3. To identify for human review records likely to be duplicates

During the summer of 1989, the author conducted a study of how these goals could be accomplished. Programs to implement the findings of the feasibility study were written during the winter of 1990 and were run against the database during April 1990–March 1991.

Basic Procedure

An early decision was made to perform as much of the cleanup and deduplication as possible on PCs, with the assumption that this would be more expeditious than developing programs to run on the CDC mainframe in Ottawa. In fact, it turned out that the entire project was done on PCs.

However, the required time would have been too great and disk space (for work files) too large to process the 140-megabyte database in one batch. Each record required about two seconds of processing just for cleanup, and the time to match records approximately squares as the number of records doubles. Therefore, records were copied from the mainframe onto diskettes in pools of about five thousand records and sent from Ottawa to Systems Planning in Northern California. After cleanup and duplicate matching, the sets of matched duplicates were sent to GCI in Southern California, where the editors decided which sets were true duplicates. The results were sent back to Systems Planning for merging, after which the complete file was sent to CHIN for uploading to the online file, replacing the earlier versions of the records. Because of the size of the data sets, Federal Express was used for all transfers.

The remainder of this paper reports on the cleanup and deduplication work in detail.

CLEANUP

Error Types

During the feasibility study for this project, three error types were defined:

1. Type 1 errors are those that a computer program can recognize and correct. For example, diacritical marks in five styles that have to be made uniform are type 1 errors; other type 1 errors are missing or incorrect punctuation, alphabetic data in numeric fields, and data in incorrect fields.
2. Type 2 errors are those that a program can recognize but a human being must correct; the corrections may or may not require the human editor to have the source document in hand. A common example is a missing mandatory field.
3. Type 3 errors are those that a program can neither recognize nor correct, such as a title incorrectly cited. These were not much studied, since the focus of this project was machine correction.

The programs were designed to correct type 1 errors and to put messages in records with type 2 errors.

At the outset of program development, sixty-five classes of errors of types 1 and 2 were identified. By the time the runs were finished, this number doubled as data-management staff discovered additional classes of errors during regular database maintenance activities.

Results of Cleanup

More than 2.1 million fields were replaced by the programs to correct type 1 errors, for an average of about 18 fields per record. In addition, 49 percent of the records were found to contain type 2 errors and were flagged for later human correction. These programs have made a large difference in the quality of BCIN. Not only has the appearance of the file been im-

proved, but corrections to indexed fields mean that retrieval has improved. An informal study showed that queries now retrieve 20 percent more hits than the same queries did before the project began, despite removal of duplicate records.

REMOVAL OF DUPLICATE RECORDS

Discussion of Deduplication Theory and Literature Review

The planning for any deduplication project must address the following topics:

1. The goal of duplicate checking
2. The question of whether a computer or an editor determines which records are duplicates
3. The processing algorithm
4. The selection of fields to use for duplicate checking
5. The preparation of the match keys
6. The determination of what is considered a duplicate
7. The question of what to do with duplicates once identified
8. The need for testing

Each of these topics is discussed below.

The Goal of Duplicate Checking

It is important to realize that no deduplication project can entirely eliminate duplicates without the risk of also eliminating valid records. A balance must be found between a method that is too loose and one that is too tight. One cannot always distinguish between errors and legitimate differences in cataloging practices.

The literature suggests that there are two basic approaches to duplicate checking, depending on the project goals. One approach, which I call the loose method, seeks to match records for human review; since the reviewers will be making the final determination, more rather than fewer matches are desired, and thus records with a lower degree of similarity will be matched. The tight method is for projects in which human review is impossible; records must match very closely because they will be merged automatically.

Generally, the loose method either uses a few fields in the comparison of records or uses more fields but requires only a few of them to match. The tight method generally uses more fields and requires a high degree of matching.

Should the Computer Remove the Duplicates?

Because of the many uncertainties in cataloging theory and practice, it is impossible to design a deduplication algorithm that will remove all the duplicates without removing some legitimate records. Since human beings are far better at perceiving patterns and nuances than computers, and computers are better at comparing large numbers of records to suggest possible duplicates, a strategy should be used that builds on these different abilities.

The Processing Algorithm

A two-pass matching process conserves computer time.[1] The first pass inspects the entire database and groups records into "pools" that have some possibility of being duplicates. The second pass examines each pool with a more precise algorithm to find duplicates. The advantage of this two-pass process is that it saves the computer's having to examine minutely each record against all others. Since the precise algorithm is used on only a few hundred or a few thousand records at a time, the system is less burdened. The comparison may even be able to take place in memory.

The fields used to form the pools are related to, but may be different from, the fields used for the final duplicate checking. These fields can be determined only by studying the file.

Selection of Fields to Use for Duplicate Checking

Like the fields used to build the pools, the fields used for duplicate checking can be determined only by studying the file. It is not a matter of choosing just fields, but possibly parts of fields (first and last words of titles, for example).

The fields used in previous work have varied widely (see table 1). In general, the literature is divided between using few data elements with human review[1,2] and using many to avoid the need for human review.[3-5]

Table 1. Fields Used for Deduping in Selected Projects

	BNB, OCLC (Goyal)	Citation Databases (Onorato & Bianchi)	Citation Databases (Slach)	Citation Databases (Yannakoudakis et al.)	Illinois Union Cat (Williams & MacLaury)	MELVYL (Coyle)	OCLC (Coyle)	OCLC (Hickey & Rypka)	RLIN (Coyle)[11]	Toxline (Giles et al., 1976a)	Toxline (Giles et al., 1976b)	WLN (Coyle)
Author		x	x	x	x	x	x	x		x	x	x
Bibliographic level							x					x
CODEN										x		
Country of publication						x						x
Date	x	x	x		x	x	x	x	x	x	x	x
Edition	x					x		x	x			
ISBN						x		x				
Journal title										x	x	
Language	x											
LCCN						x	x	x	x			x
Length of title	x											
Literature form							x	x				x
Pagination			x		x	x	x	x		x	x	
Place of publication							x	x	x			
Publisher	x					x	x	x	x			
Report number												
Reproduction type						x	x	x				
Series title								x				
SuDocs number								x				
Title	x	x		x	x	x	x	x	x	x	x	x
Volume number	x									x	x	

Preparation of the Match Keys

Most databases contain a variety of data errors and/or inconsistencies that can affect duplicate checking. Errors range from typos to fields entirely missing. Inconsistencies usually result from the differing judgment or standards of the catalogers. The errors and inconsistencies must be eliminated as much as possible to get the best duplicate matching.

Therefore it is standard practice to perform duplicate matching using match keys—normalized strings built from the fields chosen for deduplicating the file. The normalization usually eliminates spacing, punctuation, special characters, and diacriticals and converts all letters to uppercase.

Further considerations in this area concern ways to reduce the effects of minor typos, missing articles, and slight variations in wording. Some of these techniques are truncation, keywording, hashing,[5] finding the Hamming distance between Harrisoned strings,[7] Hamming and Harrissoning, soundex and similar techniques,[3,4,6,7] and the Universal Standard Book Code (USBC).[8,9] The choice of which to use depends on the file; there are no universal rules. The USBC as described by Goyal particularly seems to depend on clean data, especially on a clean title. Yannakoudakis et al. report on experiments with the USBC in matching records from various sources.

What Is Considered a Duplicate?

When the machine is matching duplicates, what is its definition of *duplicate* to be? If a slightly different pagination exists in two records with the same author, article title, journal title, and year, it is likely that one pagination is a mistake, and the editor should be shown the match. However, in the same case, if the article titles are different, then maybe the author wrote two different articles in the same year for the same journal—the likelihood of duplication here is much smaller. Considerations such as these can be used to assign weights to each field used for matching.

The resulting scores for a pair of records are then compared against threshold values to determine whether to consider the pair duplicates.

What to Do with Duplicates Once Identified

There are three basic choices for merging records determined to be duplicates:

1. One record is chosen as the master record and the others are deleted.
2. All records are kept but clustered with a master record.
3. One record is chosen as the master record and variant fields from the duplicates are added to the master.

Apart from number 2, which may be difficult to implement in some systems, the decision is a matter of policy.

The Need for Testing

The need for testing in designing a duplication project cannot be overemphasized. Different techniques work for different databases. Normally this is an iterative process: a method is proposed and tested on a sample of records from the main database; the results of the test suggest modifications to the method. This cycle is repeated until an acceptable level of success is achieved.

Deduplicating BCIN

The task of identifying duplicate records in BCIN was shaped by the following considerations:

- Because BCIN was initially formed by merging databases from contributors who did not use a single cataloging standard such as *AACR2*, BCIN has a low uniformity even when compared with other union catalog projects. Although the cleanup was performed before duplicate matching in order to increase uniformity as much as possible, the fact remains that differences in cataloging standards generally result in type 3 errors (those a program cannot detect).
- BCIN contains catalog records for serials, serial analytics (articles), monographs, monograph analytics (chapters), unpublished reports, and audiovisual material. The programs had to find duplicates among all of these but serials (the subject of a separate GCI project). Because of variations in cataloging, records for these material types could not be processed separately but had to be compared against one another.
- In a specialized database like BCIN,

certain authors, title words, journal titles, and conference names occur with high frequency. With missing volume or issue numbers, missing pagination, and incorrectly formatted dates, it is often impossible for the programs to distinguish true duplication from coincidental similarity.

- Processing on a PC was mandated to reduce the work load on CHIN.

These considerations resulted in a decision to have the computer identify potential duplicates, but to leave the final decision about duplicates to human editors. Thus this project is an example of the loose method.

The decision to perform duplicate matching on IBM-compatible PCs mandated a pooling method, although regardless of equipment it may have been infeasible to compare 140,000 records with one another; this would have required ten billion comparisons.

The database was divided into twenty-three pools of about five thousand records each. Records were assigned to pools based on the date of publication, since that field was found to be the most accurate and uniform among those useful for deduplication. However, choosing that pooling method meant that if a record had a missing or incorrect data field, the record would not be included in the comparisons with other records for that bibliographic item. To compensate for this, after the entire database is deduplicated, new pools will be created based on the author fields, and the database will be deduplicated again.

Work Flow

Each pool was created only when needed. This required only a small part of the database to be frozen at any time. Each pool was created by CHIN using normal BASIS database commands and sent on diskettes to Systems Planning. Systems Planning then ran the cleanup program (Record Validation) and the programs to match duplicates (key generation, key matching, and chain building).

The matching of records for duplicates was done by means of a match key. Each record had a 120-byte match key created for it, with fixed-length segments based on data in the record. These match keys were then compared against all other match keys in the same pool to identify possible sets of duplicates.

The results of the matching operation were sent to GCI for removal of duplicates. Human editors compared the matched records and made the final decisions.

If there were only two records in a set of possible duplicates, they were compared using a display program (View/Edit) written for this project; View/Edit allowed the editors to view pairs of records with their common fields displayed in pairs so the records appeared to be shuffled together. Each record was displayed in a distinctive color to make identification easy. The editors could choose the master record of the pair into which the other would be merged if they were true duplicates; editors could also indicate field substitutions if fields in the master record were less preferable than in the other. As many as fifty pairs of records per hour could be reviewed with this program, which also supported text editing of the merged record.

If there were more than two records in the set of possible duplicates (as many as fifty existed in some sets), the editors printed out the records in each set to compare. They keyed their decisions into a file using simple codes to control the merging of the sets.

The results of deduplication were returned to Systems Planning, which merged and reformulated the pool and sent it to CHIN on diskettes for loading. Since BCIN is a union list, all site-specific holdings data were merged into the master records.

The Chain-Building program deserves a special note. (This was the program that assembled sets of possible duplicates prior to editorial review.) The reason for assembling sets may be illustrated by the following example. If record A is similar to B, and B to C, and C to D and E, and D to F, and E to G, the editor needs to see all seven of these records together, even though records A and G may have little in common. Since these records are all those with similar data, by logic this also means that a record cannot appear in more than one set. Thus the chain-building program examined the

pairs of record numbers output by the key-matching program and linked together all pairs with similar data.

Choice of Fields

A review of earlier studies resulted in this compilation of fields used for deduplication:

Author
Bibliographic level
CODEN
Country of publication
Date
Edition
ISBN issue number
Journal title
Language
LCCN
Length of title
Literature form
Pagination
Patent number
Place of publication
Publisher
Record type
Report number
Reproduction type
Series title
SuDocs number
Title
Volume number

For BCIN deduplication, it was considered important to find fields with the following characteristics:

1. Frequency. There is no point in using a field that only a few records contain. This ruled out (from the list above) CODEN, country of publication, edition, ISBN, language, patent number, place of publication, and report number. Record type was of no value because all the records were of the same type.

2. Consistency. If the data in the field are inconsistent, they provide a less certain means for matching. Machine approaches to reducing inconsistency can be used, but some kinds of inconsistency can be resolved only by people. For BCIN, the following fields would have required human correction and thus were ruled out: literature form and publisher. Length of title was not used for the same reason.

3. Simplicity of programming and processing. Simplicity of programming argued for having all records use the same fields, rather than choosing fields based on bibliographic level or format (although of course not all records contained all fields). This ideal would have been dropped if too restrictive. Processing simplicity was achieved by using fields requiring little processing. Author and title are not such fields but of course were mandatory for deduplication. In general, short fields and numeric fields were preferred.

The publisher field—even had it been uniform—was not considered useful because if two monographs match on author, title, and year, they are probably either (1) true duplicates and have the same publisher or (2) in a monographic series so the publisher will not help discriminate between them. (The volume number was not reliable enough to distinguish between these two cases.)

In the end, the special characteristics of BCIN mandated that the fields used were:

Bibliographic level (monograph or analytic)
Personal or corporate author
Analytic title (title of article or chapter)
Title of main work (monograph or serial)
Series title
Date of publication
Volume number
Issue number
Pagination

Algorithm Development and Testing

Part of the feasibility study was a literature survey and study of a 1 percent sample of the database to determine the best way to identify duplicate records and what errors would interfere with identification. The proposed method was then tested by writing a program and running through it about four thousand BCIN records in several groups known to be problematic; these tests indicated slight improvements to the algorithm.

In processing fields for inclusion into the match keys, the rule was to try the simplest approach until it proved not to work in testing. The data were converted to normalized truncated keywords, but the more

complex techniques such as Harrisoning were felt to apply only to minor variations in cataloging; the larger variations in BCIN would be overcome only by using a lower threshold.

Duplicate-Matching Algorithm

Pools. The initial step of extracting a pool of records for processing was based on the year of publication, which in BCIN was the most reliable data element. The date-of-publication field was used if it existed. If not, a second field, called the date of issue, was used. In either case, only the last two bytes of the year were taken. If the field contained no numerics, the record was put into a catchall pool that was processed last.

When the pool was processed, each record had a fixed-length match key created for it from the fields discussed above.

Each field went into its own fixed-length segment. Data were left-justified within each segment. Segments were separated by a blank space for readability during testing.

If a field was missing from a record, blanks were put into its segment of the match key.

Record Number. The record number (a one- to six-byte number) was put into a six-byte segment of the match key for control purposes.

Volume and Issue Numbers. For each of these, up to four bytes were taken, starting with the first numeric character and ending with the first space or comma after the numerics. Leading zeros were deleted. These data occupied four-byte segments. If there were no numerics, four blanks were added to the match key.

Bibliographic Level. The first character of the field (either *M* for "monographic" or *A* for "analytic") was taken and made uppercase.

Titles. The analytic title (if any), the title of the main work (monograph or serial), and the series title (if any) were all used in matching records. The first three significant words of each title field were each placed in an eight-byte segment, for a total of nine segments. Each of the three fields was processed as follows:

1. The following punctuation was removed:

. , - = ‘ ” () [] ; : / <> !

2. Any @ and * and the character following were removed (these were font codes).

3. The title was converted to all uppercase.

4. The following stopwords were removed: &, A, ABOUT, AN, AND, AS, AT, AU, BUT, BY, D, DA, DANS, DAS, DE, DEN, DER, DES, DI, DIE, DU, EL, EN, ET, FOR, FROM, IN, L, LA, LE, LES, LOS, MIT, OF, ON, OR, POUR, S, SUR, THE, TO, V, VAN, VON, WITH, Y.

5. Diacritical marks were removed.

6. The first three words remaining were moved into the match key. If there was only one or two words, blank segments were added.

Authors. If both personal and corporate authors existed, the personal author was used. The personal author was processed as follows:

1. All spaces were removed.

2. The field was converted to all uppercase.

3. The field up to the comma was considered the surname. It had all punctuation removed (using the same list as for titles).

4. If the surname began with MAC, the A was removed.

5. If the surname was not ANON (anonymous), the first eleven bytes were added to the match key.

6. If the name was ANON, eleven blanks were added to the match key.

In earlier tests, the personal author's first initial was added as a twelfth byte, but this was later found to add nothing and occasionally prevented a valid match.

The corporate author was processed as follows:

1. All punctuation and font codes were removed (using the same lists as for titles).

2. The field was converted to all uppercase.

3. The first space was found. Up to eleven bytes of the first non-stopword (same list as for titles) were taken and added to the match key.

Pagination. The pagination field was valid only for analytics. Monograph pagination was in the collation field, which in BCIN was too erratic to permit extraction of a page number. The pagination field was truncated at the first hyphen, if any; then the field was processed just as the volume and issue number fields.

Matching Process. Each record in the pool had its match key compared against all other match keys in the pool. Matching was done on a segment-by-segment basis, so that it was known which segments matched and which didn't. Generally, each segment was compared against the corresponding segment in the other match key, but titles were treated differently. If both records were analytics, only the analytic titles were compared. If both were monographs, only the main titles were compared. If the two records were not of the same level, then the analytic, main, and series titles were compared against all three of those titles in the other match key. This was to catch cases where the same work had been cataloged twice, once as an analytic and once as a monograph, or once as a monograph and once as a monographic series. Scores were kept of the segments matched and compared against thresholds (described below).

The number of match-key comparisons required to compare n records is

$(n - 1) + (n - 2) + (n - 3) + \ldots + 1$

or

$(n^2 - n)/2$

For 5,000 records, the typical pool size, this amounts to 12,497,500 pairs of match keys to compare. When it is considered that there are up to thirty-two segment matches per pair, the true processing load becomes evident. A typical matching run of this size took four hours on a 386/25 computer. This computer had a 19ms ESDI cached hard disk; however, the process is CPU-bound, not I/O-bound. The generation of 5,000 match keys took less than ten minutes, the chain building about an hour.

Although time-consuming, this segment-by-segment and record-by-record matching was considered necessary for BCIN. The method of sorting match keys discussed by Giles et al.[3,4] is a faster alternative but requires a higher degree of data reliability. Although the pooling used for BCIN depends on an accurate date field, the other fields need not be so accurate.

Thresholds

It can be seen from the list of fields used for BCIN matching that four fields are primarily alphabetic in nature (the author field and the three title fields); furthermore, since three words were taken from each of three title fields, there was a total of ten segments of the match key that were primarily alphabetic. Four other fields were primarily numeric (volume number, issue number, pagination, and date); the bibliographic level (being one letter) was included in the numeric fields for convenience because of its simplicity.

Thus it can be seen that two identical records would match on ten alphabetic segments and five numeric segments. This score would be represented by 10,5.

Experiments were performed to determine how many fields had to match to consider a pair of records duplicates. Too low a threshold and the editors had to examine false matches; too high a threshold would mean missing duplicates.

The thresholds used changed during the project. At first, if both records were analytics, the threshold was 4,3 but in all other cases the threshold was 4,2 (this was because analytics contribute to more numeric fields). For later pools, the threshold was 3,2 if both records were either analytics or monographs, and 4,2 if they were of different types.

At thresholds just one point lower (either alphabetic or numeric), the sets of potential duplicates were double the size, and nearly all the increases were false matches.

Because all potential duplicates were to be reviewed by human editors, there was no attempt to rank matches by score. In fact, once the threshold was reached by a pair of records, no further comparison was done.

The lowness of these thresholds compared with a perfect score of 10,5 is an indication of the inconsistency of the data and the reason that automatic deduplication was not feasible.

RESULTS OF DUPLICATE REMOVAL

It is difficult to describe the percentage of potential duplicates that the editors considered true duplicates. A very high percentage of the chains contained duplicates. However, because some potential duplicates were in long chains, only 20 percent of the records sent to the editors were duplicates. In all, 8.24 percent of the records in the database were merged and removed by this process (another 13 percent were partial records merged by an earlier project).

CONCLUSIONS

Several lessons from the BCIN project are worth noting. First, no database is so inconsistent that a procedure cannot be designed to improve it, such as the correction of 2.1 million fields in BCIN.

In addition to the errors corrected by the programs, 59,373 records had type 2 errors (those the computer could detect but not correct); these records were flagged for later correction. In addition, there were of course type 3 errors the computer could not even detect. Thus another lesson is that programs are clearly not a complete solution to the problems of cleaning up databases.

Third, an inconsistent database requires human editors to determine which records are duplicates. Although this is costly, this project bore out Onorato and Bianchi's contention that automatic duplicate removal would be so complex as to be even more costly than human review.[10]

The addition to each BCIN record of its pool number and date of processing provided an audit trail of great value when programming errors or additional cleanup requirements were noted. Furthermore, each duplicate record merged was noted in its master record by number and date. Both these ideas are recommended.

Finally, never underestimate the time it takes for a project of this kind. Although at the end one pool per week was being processed, the first pool took months, and several were processed before smoothness was obtained.

ACKNOWLEDGMENTS

The author wishes to thank Jacqueline Zak, GCI's data manager and project manager of the BCIN cleanup and deduplication; her gimlet-eyed assistant Linda Kincheloe; and their manager Kathleen McDonnell for their unfailing support and patience throughout the project. Their review of this manuscript was also invaluable, although any remaining mistakes are the author's own.

REFERENCES

1. Martha E. Williams and Keith D. MacLaury, "Automatic Merging of Monographic Data Bases—Identification of Duplicate Records in Multiple Files: The IUCS Scheme," *Journal of Library Automation* 12, no.2:156–68 (June 1979).
2. June E. Slach, "Detection and Elimination of Duplicates from Multidatabase Searches," *Bulletin of the Medical Library Association* 73, no.3:235–37 (July 1985).
3. C. A. Giles and others, *A Computerized Scheme for Duplicate Checking of Bibliographic Data Bases* (NTIS CRNL/CSD-5, Aug. 1976).
4. C. A. Giles and others, *An Experiment in Computer-Assisted Duplicate Checking* (NTIS CONF-76-1011-2, n.d.).
5. Thomas B. Hickey and David J. Rypka, "Automatic Detection of Duplicate Monographic Records," *Journal of Library Automation* 12, no.2:125–42 (June 1979).
6. James L. Dolby, "An Algorithm for Variable-Length Proper-Name Compression," *Journal of Library Automation* 3, no.4:257–75 (Dec. 1970).
7. William R. Nugent, "Compression Word Coding Techniques for Information Retrieval," *Journal of Library Automation* 1, no.4:250–60 (Dec. 1968).
8. Pankaj Goyal, "Duplicate Record Identification in Bibliographic Databases," *Information Systems* 12, no.3:239–42 (1987).
9. E. J. Yannakoudakis and others, "Matching of Citations Between Non-Standardized Databases," *Journal of the American Society for Information Science* 41, no.8:599–610 (1990).
10. E. S. Onorato and G. Bianchi, "Automatic Identification of Duplicates After Multidatabase Online Searching," *Online Review* 5, no.6:445–51 (Dec. 1981).
11. Karen Coyle, "Record Matching: A Discussion," *Information Technology and Libraries* 4:57–59 (Mar. 1985). ■■

Network-Based Electronic Serials

Charles W. Bailey, Jr.

New forms of scholarly communication are evolving on international computer networks such as BITNET and Internet. Scholars are exchanging information on a daily basis via computer conferences, personal e-mail, and file transfers. Electronic serials are being distributed on networks, often at no charge to the subscriber. Electronic newsletters provide timely information about current topics of interest. Electronic journals, which are often refereed, provide scholarly articles, columns, and reviews. Utilizing computer networks, scholars have become electronic publishers, creating an alternative publication system. Electronic serials hold great promise, but a variety of problems currently limit their effectiveness. Given the serials pricing crisis, librarians should encourage the development of network-based electronic serials.

Librarians have been hearing about electronic serials for a long time. There are a growing number of full-text serials on database vendor systems like Dialog; however, they are usually derived from print serials, and they are available on a pay-per-use basis. These full-text databases provide improved access to serials information, but libraries cannot purchase and own them. Some full-text serials databases are available on CD-ROM. These databases are typically licensed at a flat annual fee. Outright ownership of these databases is usually not a purchase option.

As a result, commercial firms own the information in these full-text electronic serials, and, one way or another, we rent it. Is access versus ownership a problem? It depends on several factors. If we pay as we go, how frequently must the information be used? How much will it cost? How will it increase or decrease the cost of accessing serials information? How rapidly will these costs rise over time? How easily can we obtain the print equivalents of these electronic serials if the vendor discontinues the electronic version or we can no longer afford it? How sure are we that information that is solely in electronic form will be preserved?

As we ask ourselves these questions, we must remember that no library will be able to provide these materials to us via interlibrary loan. If we are lucky, there may be a few alternative commercial suppliers for full-text electronic serials, but we will usually be in an information monopoly situation. We will have rapid access to selected serials information using powerful searching techniques, but we will have paid for this improved access by sacrificing ownership.

Given these issues, it is fair to ask the question: Where is the promised revolution in scholarly communication that electronic serials publishing was supposed to bring?[1]

THE NET

To find the answer to this question, we must turn our attention from commercial electronic information systems to noncommercial international computer networks.

Charles W. Bailey, Jr., is Assistant Director, Systems, University Libraries, University of Houston. This paper was delivered at the Sixth Annual Conference of the North American Serials Interest Group, June 15, 1991.

BITNET is a sprawling computer network that links more than three thousand computers in thirty-five countries.[2] BITNET has gateways that connect it with other large, noncommercial networks, like Internet, as well as with commercial networks, like CompuServe. BITNET supports person-to-person e-mail and file transfers in addition to thousands of computer conferences on a wide range of topics.

Internet, an international network of networks, is growing at such a rapid rate that estimating its size is more difficult; however, it may connect at least four hundred networks and close to a half-million computers.[3] Internet offers similar services to BITNET, plus remote login to computers via the TELNET protocol.

These two networks, plus numerous other networks, form a burgeoning, complex, worldwide electronic communication system that is irrevocably changing the process of scholarly communication. This collective entity, which I will simply call the Net, will force us to rethink many of our print-based assumptions about the selection, collection, organization, provision, and preservation of knowledge.

The information packages that we are used to—primarily books and journals—have many characteristics that are artifacts of print technology. For example, we expect that a book or journal will have a familiar overall structure and layout, with features such as numbered pages. We know that it will be an immutable physical object, and we accept certain distribution and storage strategies that are necessitated by this fact, such as the grouping of a number of articles into a single journal issue. Normally, we anticipate that books and journals will be formal communication tools, with care given to their composition. Informal messages are usually reserved for one-on-one or small group communication activities.

Electronic communication on the Net has created new types of information packages, such as computer conferences and electronic serials, that are not necessarily bound by print conventions. As technology advances and mimicry of print declines, these information packages will become increasingly unfamiliar, and we will see the emergence of complex, information-rich multimedia computer systems on the Net.[4,5]

THE EMERGENCE OF NETWORK-BASED ELECTRONIC SERIALS

During the last few years, a small group of pioneering publishers on the Net has been distributing electronic journals, magazines, and newsletters that are not derived from print serials. Publishers send these electronic serials to subscribers via e-mail messages or file transfers. Usually, they are free of charge, and there are typically no license agreements.

When this small group of pioneering publishers looked at the Net, they saw an opportunity to accelerate the evolution of scholarly electronic communication. Often using software designed to support computer conferences, they set themselves up as electronic publishers. Some of these electronic publishers are producing serials that are similar to familiar print publications; others are inventing new forms of serials.

ARE COMPUTER CONFERENCES SERIALS?

Before we discuss the efforts of these electronic serials publishers, I'd like to pose a question: Are computer conferences serials?

AACR2 defines a serial as "a publication in any medium issued in successive parts bearing numerical or chronological designations and intended to be continued indefinitely."[6] You will note that this definition does not mention the type of information contained in serials, nor does it specify that serials must be edited works.

Each e-mail message sent out on a computer conference is identified as originating from that conference, and it could be viewed as being a "successive part" of the conference. All messages have a date and time stamp, and, thus, have a "chronological designation." Most computer conferences are intended to be continued indefinitely.

Let's look beyond basic compliance with the definition of a serial. Messages have identified authors. Although many messages are short, some can be quite lengthy, at times exceeding five hundred lines. Some computer conferences are "edited" by a conference moderator. At a basic level, this means that the moderator screens all incoming messages and sends out only those that are relevant to the conference. At a more ad-

vanced level, the moderator may assemble messages into formal "digests" and send these composite messages to conference participants. Some moderated conferences, like the Humanist, issue messages that have volume and number designations.

Computer conferences on BITNET and other networks are increasingly vital channels of scholarly communication. Are they serials? Quite possibly.[7]

If computer conferences are serials, librarians need to ponder the issue of how important these conferences are and whether their proceedings should be preserved. We cannot assume that computer centers will treat these information resources as libraries would. The selection, acquisition, organization, provision, and preservation of knowledge is the domain of libraries, not computer centers.

EXAMPLES OF NETWORK-BASED ELECTRONIC SERIALS

What electronic serials are available on the Net? I'll provide a few representative examples of scholarly electronic newsletters and journals. There are also network equivalents of popular and special-interest magazines. Currently, the task of identifying network-based electronic serials is burdensome. However, in the near future, the Office of Scientific and Academic Publishing of the Association of Research Libraries will publish an authoritative list of network-based electronic serials and computer conferences.[8]

Many electronic serials on the Net make use of the revised LISTSERV software, which was written by Eric Thomas.[9] The LISTSERV software is usually referred to as the "list server."

ELECTRONIC NEWSLETTERS

There are a number of electronic newsletters on the Net. These publications usually have small editorial staffs. In addition to short articles, news items, and editorial commentary, electronic newsletters typically contain a variety of reader-submitted material, such as brief comments on current issues, conference announcements, job listings, and other short information items. Issues are usually sent out as e-mail messages. These newsletters are typically issued on an irregular basis. Back issues may be stored as files on a network computer where users can retrieve them directly; however, in some cases, back issues must be requested from editorial staff. Normally, there are no subscription fees.

- *ALCTS Network News*, a publication of the ALA's Association for Library Collections and Technical Services division, keeps subscribers informed about the activities of that organization.
- *CCNEWS*, an information resource for academic computing newsletter editors that is published by EDUCOM, comes in two separate sections: (1) a newsletter and (2) a collection of abstracts for article files that have recently been added to the *CCNEWS* list server.
- *Health Info-Com Network Newsletter* is about medical topics. In addition to typical newsletter material, scholarly articles appear in this publication.
- The *Newsletter on Serials Pricing Issues* is about the serials pricing crisis and related serials issues.[10]
- *The Online Journal of Distance Education and Communication* is about distance education and communications, telecommunications in education, and electronic cross-cultural communication.
- *Public-Access Computer Systems News*, a publication of the University of Houston Libraries, is about computer systems that libraries make available to their patrons.

ELECTRONIC JOURNALS

There are a growing number of electronic journals on BITNET, most of them founded in 1990–91. These journals typically have several people on the editorial staff and a formal editorial board. Many of these journals are refereed. They often mirror print journals, containing editorials, scholarly articles, columns, and reviews. Issues can be composed of a single article or multiple articles. Single-article issues are often sent as e-mail messages. Multiple-article issues may be announced via an e-mail message, with users retrieving article files of interest based on this message. Or, if the journal has relatively short issues, it may be distributed as an e-mail message. Some attempt regular publication cycles; many are irregular. Issue or article files may be archived on a network computer so

that users can retrieve them as needed, or they may be available from editorial staff on request. Usually, the journal is not indexed in conventional sources. There is normally no subscription charge for issues distributed on the Net. A subscription fee may be charged for an alternative distribution format, such as floppy disk.

- *EJournal* is about electronic texts and computer networks.[11] It is anticipated that this journal will experiment with archived reader responses to articles, author revisions of articles, and article retractions.
- *The Electronic Journal of Communication/La Revue Electronique de Communication* is a bilingual journal about the study of communication theory, research, practice, and policy.[12] The journal is distributed via Comserve at Rensselaer Polytechnic Institute.
- The *Journal of the International Academy of Hospitality Research* is about hospitality and tourism issues.[13] It is published by the Scholarly Communications Project of Virginia Polytechnic Institute and State University. Unlike other BITNET electronic serials, this journal charges users for subscriptions, with prices ranging from $10 to $30.
- *New Horizons in Adult Education*, a publication of the Syracuse University Kellogg Project, is about adult education.[14] Established in 1987, it is the oldest BITNET journal that I know of.
- *Postmodern Culture* is about contemporary literature, theory, and culture.[15] There is a related computer conference, which is called PMC-Talk. For a small subscription fee ($15–$30), the journal is available in floppy disk or microfiche format.
- *PSYCOLOQUY* contains peer-reviewed "scholarly skywriting"— brief statements of new ideas or findings about psychology and related fields.[16] It also functions as an electronic newsletter. *PSYCOLOQUY* is sponsored by the Science Directorate of the American Psychological Association.
- *The Public-Access Computer Systems Review*, a publication of the University of Houston Libraries, is on the same topic as *Public-Access Computer Systems News*.[17] Both publications are associated with a computer conference known as PACS-L.[18]

THE FUTURE OF ELECTRONIC SERIALS

How will network-based electronic journals fare in the future?[19] I do not anticipate that electronic serials will displace print serials in the next ten years; however, they will become an increasingly important parallel form of scholarly communication. It is possible that a significant nonprofit serial publication system will emerge from the efforts of network-based electronic serials publishers. This system is likely to be characterized by low or no subscription fees and the retention of intellectual property rights by authors. Given the grim realities of the current serials pricing crisis and the apparent dearth of viable solutions to this crisis, librarians have a vested interest in trying to make this vision a reality.

Numerous problems need to be overcome; however, many of these problems appear solvable given the dedication of adequate resources to this task. I will now discuss some key problems related to network-based electronic serials.

First, electronic serials are often distributed as ASCII text files. This distribution strategy enables users to manipulate files easily, and it minimizes data-transmission overhead, but it significantly limits the kind of information that can be represented (e.g., no color, foreign characters, illustrations, or mathematical notation). PostScript, TeX, and other software tools provide limited solutions to this problem, but no easy-to-use, ubiquitous solution currently exists.

Second, many network users have fairly limited storage capacity in their computer accounts. Last year, one new electronic journal overwhelmed many of its subscribers' accounts by sending them more than two hundred pages of information at one time.

Third, users may not understand the mechanics of network e-mail and file transfers, the operation of useful mainframe software, or downloading procedures.

Fourth, existing tools for creating, distributing, and utilizing network-based electronic serials are in an early stage of development, and they lack many desirable technical capabilities.

Fifth, as electronic serials proliferate, lengthy file transfers will become more common, potentially creating network performance problems. Bottlenecks may occur on network links that operate at relatively low speeds. This will become particularly problematic as electronic serials transcend the ASCII format and their files become significantly larger.

Sixth, access to networks such as BITNET and Internet is mainly limited to academics and researchers. The electronic highway is there, but not everyone can get on it. The linkage of commercial networks like CompuServe and the Well to Internet has helped solve this problem. However, users of commercial networks must pay to use these services, while many users of BITNET and Internet have subsidized access to network services. Many potential foreign readers, especially those in the Third World, may not have access to U.S.-based networks at all.

Seventh, getting information about network-based electronic serials is currently a daunting task. If one knows where to look, there are a few directory files; however, they may not be up-to-date. The Association of Research Libraries' (ARL) directory will make a major contribution to solving this problem.[20] The problem is made worse by the fact that few electronic serials are included in conventional printed indexes and abstracts. Once you find an electronic serial, determining the contents and availability of back issues can be problematic.

Eighth, the publication of electronic serials is still a somewhat subterranean activity. Mostly, they are published by academics, but this effort may not be recognized as an official university activity. Electronic serials are mainly the work of a few dedicated, volunteer editors, not well-staffed institutionalized "electronic presses." This lack of institutional structure and support has a variety of problems associated with it.

Ninth, although electronic serial files are currently being archived at specific network nodes, there is no guarantee that computer centers will indefinitely preserve these files, especially when no high-level institutional commitment has been made to do so.

Tenth, the application of existing intellectual property laws to the subtleties of electronic publishing is not all that straightforward.[21] The U.S. Register of Copyrights is granting the copyright applications of electronic serials publishers; however, until there are more legal precedents, this area of the law remains somewhat hazy.

Eleventh, there is the issue of acceptance of electronic serials. Will tenure committees accept publication of an article in an electronic journal as being equivalent to publication in a similar print journal, even if the journal is refereed? Will the majority of scholars want to publish in journals that are neither indexed in conventional sources nor collected by libraries?

Finally, existing network-based electronic serials have been able to develop because a large number of users have subsidized access to networks like BITNET and Internet. If the proposed National Research and Education Network (NREN) that may replace these networks is operated by commercial firms, network economics may change, and this may adversely affect user access to both electronic serials and computer conferences.

CONCLUSION

Despite these problems, a growing number of useful journals, magazines, and newsletters are being produced for and distributed to network users. With a minimum investment of resources, professional societies, university presses, individual scholars, and librarians can utilize existing software to establish network-based electronic serials.

As I indicated earlier, network-based electronic serials are unlikely to replace printed serials from conventional publishers in the foreseeable future; however, they can provide an alternative source of scholarly information. Although there is nothing inherent in network-based publishing that mandates that electronic serials be made available free or at low cost, it is possible to do so given the radically different economics of these publications and printed serials. With future improvements in printing technology and information standards, there could be a dramatic leap in the reproduction quality of locally printed electronic serials.[22] For practical purposes, printed copies of electronic seri-

als made in the future may be indistinguishable from their conventional counterparts.

Librarians can play an important role in determining the future of network-based electronic serials. We can construct printed or computer-based tools that will help our users to identify and access network-based electronic serials. We can collect, provide local access to, and preserve these electronic serials. We can help our campuses establish units to publish high-quality electronic serials or do so on our own. We can promote the development of new standards that will improve the storage, distribution, display, and printing of network-based electronic serials. We can lobby for the establishment of a high-performance, government-subsidized National Research and Education Network.

If we do these things, network-based electronic serials may become a significant alternative source of low-cost scholarly information by the end of this decade. If not, network-based electronic serials are likely to evolve more slowly, and the serials pricing crisis is likely to continue unabated.

REFERENCES AND NOTES

1. For a discussion of electronic journals that are not distributed on the Net, see: Anne B. Piternick, "Serials and New Technology: The State of the 'Electronic Journal,'" *Canadian Library Journal* 46:93–97 (Apr. 1989).
2. Tracy L. LaQuey, ed., *The User's Directory of Computer Networks* (Bedford, Mass.: Digital Pr., 1990), p.2.
3. John S. Quarterman, *The Matrix: Computer Networks and Conferencing Systems Worldwide* (Bedford, Mass.: Digital Pr., 1990), p.278.
4. Charles W. Bailey, Jr., "Intelligent Multimedia Computer Systems: Emerging Information Resources in the Network Environment," *Library Hi Tech* 8, no.1:29–41 (1990).
5. Stephen Bulick, "Future Prospects for Network-Based Multimedia Information Retrieval," *The Electronic Library* 8:88–99 (Apr. 1990).
6. Michael Gorman and Paul W. Winkler, eds., *Anglo-American Cataloguing Rules*, 2d ed. (Chicago: American Library Assn., 1978), p.570.
7. Ann Okerson also suggests that computer conferences are serials. See: Ann Okerson, "The Electronic Journal: What, Whence, and When?," *The Public-Access Computer Systems Review* 2, no.1:18–19 (1991). (To retrieve an article on the Net, send the e-mail message "GET OKERSON PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
8. The *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* was published in July 1991. It is available from: Office of Scientific and Academic Publishing, Association of Research Libraries, 1527 New Hampshire Ave., N.W., Washington, DC 20036.
9. For a detailed discussion of the revised LISTSERV software, see: Diane Kovacs and others, "How to Start and Manage a BITNET LISTSERV Discussion Group: A Beginner's Guide," *The Public-Access Computer Systems Review* 2, no.1:128–43 (1991). (To retrieve an article on the Net, send the e-mail message "GET KOVACS PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
10. Marcia Tuttle, "The *Newsletter on Serials Pricing* Issues," *The Public-Access Computer Systems Review* 2, no.1:111–27 (1991). (To retrieve an article on the Net, send the e-mail message "GET TUTTLE PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
11. Edward M. Jennings, "*EJournal*: An Account of the First Two Years," *The Public-Access Computer Systems Review* 2, no.1:91–110 (1991). (To retrieve an article on the Net, send the e-mail message "GET JENNINGS PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
12. Teresa M. Harrison and others, "Online Journals: Disciplinary Designs for Electronic Scholarship," *The Public-Access Computer Systems Review* 2, no.1:25–38 (1991). (To retrieve an article on the Net, send the e-mail message "GET HARRISON PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
13. Lon Savage, "The Journal of the International Academy of Hospitality Research," *The Public-Access Computer Systems Review* 2, no.1:54–66 (1991). (To retrieve an article on the Net, send the e-mail message "GET SAVAGE PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
14. Jane Hugo and Linda Newell, "New Hori-

zons in Adult Education: The First Five Years (1987–1991)," *The Public-Access Computer Systems Review* 2, no.1:77–90 (1991). (To retrieve an article on the Net, send the e-mail message "GET HUGO PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

15. Eyal Amiran and John Unsworth, "Postmodern Culture: Publishing in the Electronic Medium," *The Public-Access Computer Systems Review* 2, no.1:67–76 (1991). (To retrieve an article on the Net, send the e-mail message "GET AMIRAN PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
16. Stevan Harnad, "Post-Gutenburg Galaxy: The Fourth Revolution in the Means of Production of Knowledge," *The Public-Access Computer Systems Review* 2, no.1:39–53 (1991). (To retrieve an article on the Net, send the e-mail message "GET HARNAD PRV2N1 F = MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)
17. Charles W. Bailey, Jr., "Electronic (Online) Publishing in Action . . . *The Public-Access Computer Systems Review* and Other Electronic Serials," *Online* 15:28–35 (Jan. 1991).
18. Charles W. Bailey, Jr., "The Public-Access Computer Systems Forum: A Computer Conference on BITNET," *Library Software Review* 9, no.2:71–74 (Mar.–Apr. 1990).
19. For other visions of the future of electronic serials, see: William Gardner, "The Electronic Archive: Scientific Publishing for the 1990s," *Psychological Science* 1:333–41 (Nov. 1990); Tim King, "Critical Issues for Providers of Network-Accessible Information," *EDUCOM Review* 26:29–33 (Summer 1991); Paul Metz and Paul M. Gherman, "Serials Pricing and the Role of the Electronic Journal," *College & Research Libraries* 52:315–27 (July 1991); Sharon J. Rogers and Charlene S. Hurt, "How Scholarly Communication Should Work in the 21st Century," *College & Research Libraries* 51:5–8 (Jan. 1990); "Task Force Report Looks at Future of Information Services," *Bulletin of the American Physical Society* 36:1105–1151 (Apr. 1991); and Jerome Yavarkovsky, "A University-Based Electronic Publishing Network," *EDUCOM Review* 25:14–20 (Fall 1990).
20. The *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* has helped solve the problem of identifying electronic serials, and it is the best guide to these publications that is currently available. The section of the directory that deals with electronic serials, which was complied by Michael Strangelove, is also available on the Net. To obtain it, send the following two commands on separate lines in an e-mail message to "LISTSERV@UOTTAWA: GET EJOURNL1 DIRECTRY" and "GET EJOURNL2 DIRECTRY."
21. For further discussion of intellectual property rights issues, see: Adrian W. Alexander and Julie S. Alexander, "Intellectual Property Rights and the 'Sacred Engine': Scholarly Publishing in the Electronic Age," *Advances in Library Resource Sharing* 1:176–92 (1990); Mary Kay Duggan, "Copyright of Electronic Information: Issues and Questions," *Online* 15:20–26 (May 1991); and Ann Okerson, "With Feathers: Effects of Copyright and Ownership on Scholarly Publishing," *College & Research Libraries* 52:425–38 (Sept. 1991).
22. For an interesting overview of DEC's use of on-demand printing, see: Sally Taylor, "The Potential of On-Demand Printing," *Publishers Weekly* 238:38–40 (June 7, 1991). ■■

Communications

Fall 1991 CNI Task Force

Paul Evan Peters

The following is the summary report prepared and distributed by the secretariat of the Coalition for Networked Information (CNI) on the proceedings of the fall 1991 meeting of the Task Force of the Coalition.

The coalition is a joint project of the Association of Research Libraries (ARL), CAUSE, and EDUCOM that was organized in March 1990 to promote the creation of and access to information resources in networked environments in order to enrich scholarship and to enhance intellectual productivity.

The coalition encourages you to use this summary report to provide information to other individuals in your organization or institution and to prepare articles for local newsletters or entries in local reports.

Additional information about the fall 1991 meeting of the Coalition Task Force, and the various talks presented, documents distributed, and calls issued at this meeting, can be obtained from Joan Lippincott, coalition assistant director, by phoning the coalition office at (202) 232-2466 or by sending electronic mail to joan@cni.org.

INTRODUCTION

The fall 1991 meeting of the Coalition Task Force took place on November 21–22 in Washington, D.C. Close to 300 individuals from more than 200 institutions and organizations attended the meeting. A total of 150 institutions and organizations now belong to the Coalition Task Force, and 85 percent of them were represented at this meeting. Eighteen institutions and organizations attended this meeting as new members of the task force. A special effort was made to invite representatives of the university-press community to this meeting, and more than a dozen were in attendance, including Peter Grenquist, executive director of the American Association of University Presses.

Paul Evan Peters is Director, Coalition for Networked Information, Washington, D.C.

CATALYZING THE FLOW OF NETWORKED INFORMATION

The meeting theme, "Catalyzing the Flow of Networked Information," was introduced by two speakers in the opening plenary session, Richard Katz and Czeslaw Jan Grycz, both of the University of California, Office of the President.

Katz's talk, "Academic Information Management at the Crossroads: Time Again to Review the Economics" stressed the need for information managers to come to grips with the current economic, political, and higher-education climate. He encouraged the audience to turn its attention to ways that both publishers and academic information managers could prosper by organizing a new infrastructure to lower total "life cycle" costs without hurting the publishers' ability to get a fair return on their investment. He also outlined the parameters for the "base case," a simplified economic model of the flow of information to and through academic libraries, which could be used to determine and compare the price and cost of printed and networked information more accurately.

Grycz reported on continuing progress with developing and evaluating models for the flow of information in the networked environment, a coalition priority since the meeting it sponsored on this topic in Monterey earlier this year. He called special attention to a double issue of *Serials Review*

devoted to these subjects that will appear early in 1992. Grycz recommended that the coalition codify a suite of prototypes based on specific criteria designed to provide publishers, brokers, and distributors of information with factual analyses of pricing and distribution schemes. He also recommended a study of behavioral changes evidenced by end users in the networked environment.

Further elaborating on the meeting theme, a panel examined issues related to new approaches to copyright and intellectual property in the networked information environment. Paul Evan Peters, director of the coalition, led off with an introduction of the new Rights for Electronic Access to and Delivery of Information (READI) program, that is being developed by the coalition to encourage thought and discussion about whether contract law, in the form of licenses and related agreements between creators and users of published works, can be employed within the context of copyright law to facilitate the flow of networked information. The goal of the proposed READI program is the licensing of printed and electronic materials, such as scholarly journals and books, so that they can be made available over networks by and to READI program participants.

Brian Kahin, director of the Information Infrastructure Project at the John F. Kennedy School of Government, Harvard University, focused on noncommercial publishing by exploring how the fundamental distinction between a "work" and a "copy" of that work changes when moving from the print to the networked environment. He proposed a number of core library functions for the electronic era: for instance, aggregating the use of hardware, pooling purchasing power, and providing in-person support and service.

John Garrett described some of the findings of a forthcoming report from the Corporation for National Research Initiatives (CNRI), where he is director of information resources. Observing that librarians and publishers constitute mutually dependent communities, the heart of this report develops four scenarios for digital libraries, examining structures, stake holders, and how copyright could be managed in each. The report finds that the current copyright framework is adequate for the transfer of rights in the electronic environment and that improved licensing schemes, within an international framework, are needed. He emphasized that however these issues evolve, there must be consensus, and he praised the coalition's role in bringing together the various stake holders.

Paul Gherman, university librarian at Virginia Polytechnic Institute and State University, and chair of the ARL Scholarly Communication Committee, sounded an alarm that copyright issues are control issues and asked the audience to consider who will control the production and distribution of knowledge on networks. He observed that in the networked environment, the creators, gatekeepers, disseminators, intermediaries, and archivists of scholarly information will likely all change and that this will change control structures and processes. He concluded that we will continue to have trouble imagining and building a new infrastructure for scholarly communication until all players agree upon a common set of purposes for scholarly communication.

PERSPECTIVES OF THE HUMANITIES, ARTS, AND SOCIAL SCIENCES

In a spirited and provocative address on the state and prospects for networked information resources and services in the humanities, arts, and social sciences, Douglas Greenberg, vice-president of the American Council of Learned Societies, called for information technologists and publishers to focus on problems of "access" rather than those of "excess" in the current system of scholarly communication and publication and to recognize the importance of retrospective conversion of the literature record for a wide range of scholarly and pedagogical endeavors. He also contrasted the funding fortunes and prospects of the sciences and professions with those of the humanities, arts, and social sciences to substantiate his view that in matters of technological change in academe, we get what we pay for and reward.

NREN PUBLIC-POLICY FRAMEWORK

In an address highlighting the potential economic importance of the National Research and Education Network (NREN) and electronic networking, Lewis Branscomb, Albert Pratt public service professor and director of the Science, Technology, and Public Policy Program at the John F. Kennedy School of Government at Harvard University, advocated the view that NREN public policy focus on diffusion-oriented processes rather than on mission or producer-oriented ones. Branscomb acknowledged that the public-policy framework that he recommends is not as established in the United States as those with which he contrasted it, but he noted that it is well established in Europe and Japan and that it is a most appropriate framework given the research and educational productivity objectives of the NREN. His talk had a particular sense of immediacy, as the NREN legislation was being debated on the floor of the U.S. Senate as he spoke.

OTHER MEETING ACTIVITIES

Meeting participants also attended sessions of the coalition working groups and convened or participated in synergy sessions and project briefings. The sixteen project briefings sparked particular interest and enthusiasm, as they demonstrated how far so many working prototypes of networked information resources and services projects have advanced since the spring meeting. It was widely remarked as well that all of the working groups are now clearly engaged in a variety of concrete projects that will yield tangible benefits.

CALLS FOR STATEMENTS OF INTEREST AND EXPERIENCE

Three calls for statements of interest and experience were issued at the meeting, providing a new approach to identifying and recruiting individuals and institutions who are prepared to contribute to specific coalition projects.

- The Rights for Electronic Access to and Delivery of Information (READI) Program call solicits institutions and organizations that have experience with licensing print and electronic published works and who are prepared to engage in a relatively intense series of discussions and meetings.
- The "TopNode for networked information resources, services, and tools" call seeks individuals, institutions, and organizations that are prepared to participate in the cataloging of resources for and editing of a directory of directories, catalogs, and other lists of networked information resources, services, and tools.
- The "development of a packet of information for use in formulating and addressing institutional and organizational issues arising from the emergence of a national networked information infrastructure and environment" call searches for individuals prepared to author sections in a packet of information for use in formulating and addressing institutional and organizational issues arising from the emergence of a national networked information infrastructure and environment.

COALITION INTERNET SERVER

It was announced that the new coalition Internet server, made possible by a grant from the Digital Equipment Corporation, is now operational, and Craig Summerhill, coalition systems coordinator, led a synergy session in which he presented the proposed services that this server will support. It was also announced that BRS Software Products, a division of Maxwell Online, will provide BRS/SEARCH and related support services to the coalition under an agreement signed in November. BRS Software Products is the fifth task force member to make a special contribution to the coalition.

INFORMATION POLICIES COMPILATION

"Information Policies: A Compilation of Position Statements, Principles, Statutes, and Other Pertinent Statements" was distributed to meeting attendees. Compiled by Coalition Steering Committee member Susan Brynteson of the University of Delaware, under the auspices of the Working Group on Legislation, Codes, Policies, and Practices, this compilation is an attempt to bring together in one convenient place the original text of official statements and laws related to information policy. The initial

scope of this compilation is policy statements developed by United States–based professional associations in the library and information service and technology community, supplemented by United States laws and other relevant materials. This compilation was developed in response to an increasing need for librarians, information technologists, and administrators to address situations and questions where information policy issues are at stake, often leading to development of institutional and organizational policies related to such issues. It is intended to assist such individuals by providing a resource that they can consult while formulating such policies.

1992 TASK FORCE MEETINGS

It was announced that the spring 1992 meeting of the Coalition Task Force will be held in Washington, D.C., on March 24, and 25, immediately preceding the National Net 1992 conference, and that the theme for this meeting will be "Network Navigating and Navigators." It was also announced that the fall 1992 meeting of the Coalition Task Force will be held in Washington, D.C., on November 19 and 20, and that the theme for this meeting will be "Architectures for Innovative Network Communication and Publication."

DIRECTOR'S POSTSCRIPT

The fall 1991 meeting of the Coalition Task Force was in every respect an exhilarating and gratifying event. The Coalition Steering Committee and staff came away from this meeting with a sense of accomplishment and a desire to maintain momentum. The coalition exists to assist its member institutions and organizations in their efforts to effect a historic transformation of how information is accessed and delivered, efforts that are being undertaken during extremely difficult economic and demographic times. The coalition encourages individuals in the broader networking community, as well as in its member institutions and organizations, to bring needs and issues to its attention and to contribute to its program of work. Indeed, the coalition depends on it! ■■

Special Section: First Annual VTLS Library Directors' Conference—Linking Multimedia Digital Libraries: Where We Are, Where We're Going

Introduction

Barbara L. Scheid

VTLS Inc. held its first annual VTLS Library Directors' Conference on September 19–21, 1991, at Mountain Lake Resort in Pembroke, Virginia. Eighty-two invited participants enjoyed an informal and inspiring conference. The idea for the conference arose from a suggestion made by a VTLS customer. The theme was chosen because of its national and international relevence to the library and information industries. The purpose of the conference was to provide participants materials and information they need to participate actively in the development of a new library paradigm. The conference was well received and will be an annual event.

Vinod Chachra, president of VTLS, welcomed all participants and speakers. He also gave the audience a glimpse of the evolution of computer technology and its importance to the library and information industries.

Jane Ryland, president of the Association for the Management of Information Technology in Higher Education (CAUSE), gave an overview of how a national research and education network will affect the library and information industries. She also explained the organizational structure and mission of the Coalition for Networked Information. Peter Young, director of the National Commission on Libraries and Information Science (NCLIS), reviewed the mission, programs, and projects implemented by NCLIS and explained how the planning work of NCLIS helps shape the new library paradigm. There was much discussion sparked by these two presentations, and therefore an informal question-and-answer session was held with Jane Ryland and Peter Young addressing questions from the participants. This session has been summarized for this article.

Frank R. Bridge, president of Frank R. Bridge Consulting, Inc., reviewed the library industry's adventures in a new frontier with fuzzy edges. He also gave suggestions on how to plan for library automation. Congressman Rick Boucher, U.S. Representative from the Ninth District of Virginia, gave an overview of the National High-Performance Computing Technology Act of 1991. He explained the benefits of developing a national network and high-performance computers. He also reviewed how the network will be implemented, who the players are, and how long it will take to implement.

Andriaan Theron, general manager of HP AIMS, Hewlett-Packard, Pinewood, England, presented "Advanced Image Management Solutions: Expanding Your Library's Horizons." In his presentation, Theron showed the video "1995," Hewlett Packard's version of electronic imaging in the office of the near future. If "1995" is a

Barbara L. Scheid is Public Relations Coordinator, VTLS Inc.

taste of the office of the future, every institution will need this technology in order to remain competitive.

W. H. Gerald (Jerry) Caldwell, Jr., director of Educational Systems Marketing at IBM Educational Systems in Atlanta, mesmerized the audience during his demonstration of "Ulysses." "Ulysses" is an educational program that combines text, graphics, animation, audio, full-motion video, and photographs to teach Tennyson's poem "Ulysses." The program utilizes IBM's linked-knowledge system. This educational program is a wonderful way to discover and learn about the poem. Caldwell announced that other educational programs will be available in the future.

On the second day of the conference, participants formed small groups to discuss issues such as federal funding for libraries, the impact of changing technology on the library professional, and copyright. Because of its focus, the "Electronic Libraries: The Pros and Cons of Multimedia Access Copyright Issues" session has been summarized for this article.

VTLS chose the speakers based on their expertise and did not require them to write formal papers, although some did. For this article, the formal papers presented by Vinod Chachra, Peter Young, and Rick Boucher are given verbatim. The informal presentations given by Jane Ryland and Frank Bridge are summarized. ■■

A Perspective on Linking Multimedia Digital Libraries

Vinod Chachra

The theme of this conference is "Linking Multimedia Digital Libraries." There are some key words in the conference title. The word *linking* implies networking, and the word *multimedia* in one sense implies the need for very large bandwidths. *Digital libraries* restricts the theme to computerized information access. Thus we are here to talk about the wonderful interactions between and among computers, communications, and libraries.

The perspective we have of computers has changed over the last ten to fifteen years. Our perception of communications has also undergone a change over this period. Our view of libraries needs to undergo a similar change. This is what we are here to talk about. When I first started working with computers, we thought of computers as number crunchers. Our perception of computers has evolved through six stages, from "number crunchers" to "process controllers" to "word processors" to "communication handlers" to "image processors."

Vinod Chachra is President, VTLS, Inc.

Today, computers are widely considered as "media converters," transforming image, sound, and text into their various forms such as printed material, binary bit streams, music, animation, and full motion video. Computers perceived as generalized "media converters" is the sixth stage, but there is yet a seventh stage to come that will be discussed later.

Communications technology has also gone through its own transitions. Communications once implied a collection of devices used to send messages from one location to another. Communications today implies a technology used to achieve connectivity. Connectivity is an important aspect of our society. For example, there are about 280 million people in the United States, but there are more than 5 billion telephones. Thus for every man, woman, and child in the United States, there are, on the average, 16 telephones. This statistic shows two things: that there is a great need for connectivity and most of our communications systems are location dependent. If our communications systems were location independent, each person would simply need one phone, and we would have no more than 280 million phones in the United States. There is yet another aspect of connectedness. Some of us were in Moscow in August 1991. During the attempted coup, the sound and sight of tanks in the streets of Moscow did not cause as much panic in our

hearts as did a discovery later that day that we could not get a phone line to the United States (or any other international destination). Because we felt "cut off," we made certain decisions that we would not have made had we been assured of better communications. Better communications, electronic or otherwise, results in better decision making.

Decision making is an important activity of the human mind. It is interesting to note how the development of the human mind differs substantially from that of computers. A child first learns to recognize images, then starts to work with graphics (line drawings), then moves on to text (reading), and finally works with numbers. By contrast, the development of computers has been exactly the opposite as seen in the six stages of our perception of computers. Computer development started with numbers and worked its way up to images. Images are inherently easier to work with than numbers. However, they require more computer capacity and greater communications bandwidth.

Before I describe the seventh stage of computer development, I wish to share another idea with you. This idea was first outlined by William McKeefery. He said that the impact of any given technology is best determined by the factor by which that technology multiplies human capabilities to do a given task. For example, automobiles traveling at sixty miles an hour represent only a 15-fold multiplication over walking at four miles an hour. The invention of the plow represented a 10-fold multiplier over previously known techniques for tilling the land. The introduction of chemical fertilizer was another 10-fold multiplier. The plow, along with chemical fertilizer, brought about the agricultural revolution. The invention of the steam engine was a 1,000-fold multiplier and brought in its wake the industrial revolution. Jet transportation, which has made our world smaller, at six hundred miles an hour is only a 150-fold multiplier over walking. Occasionally in human history, we come across a technology that represents a million multiplier. There are only three such technologies known to man. The first came to us in the form of communications. First by wire, and then by wireless, we learned how to send signals a million times faster than previously known techniques allowed. The second million multiplier is nuclear energy. I submit we have yet to learn how to deal with this technology. The third million multiplier is computers. With the convergence of computer and communications technologies we will for the first time have a millionfold multiplication of a million multiplier. If a 100-fold multiplication brings about the agricultural revolution, and a 1,000-fold multiplication brings about the industrial revolution, it is very difficult to predict the outcome of a millionfold multiplication of a millionfold multiplier. The task before us is to combine these two millionfold multipliers with the information content of our libraries.

The seventh stage of computer development lies in making this machine a true extension of the human mind, a device that supports human thought in every way. This stage can only be brought about by the convergence of libraries, computers, and communications. Libraries have traditionally been the warehouses of the collective knowledge of mankind. This recorded knowledge base is increasing at an alarming rate. The contents of the world's libraries represent a very substantial and potentially unmanageable resource. This conference is devoted to addressing the various issues related to bringing computers, communications, and libraries together through multimedia-linked systems. I perceive linking multimedia digital libraries an essential step in making these technologies a true extension of the human mind.

All of this promises to make for an interesting and informative day. Tomorrow, several of these speakers will lead discussion groups on topics like electronic publishing, copyright issues, federal funding, and the impact of new technologies on the professional librarian. The success of these discussions will depend largely on your participation. I urge you to take an active role in today and tomorrow. ■■

Overview of NREN and CNI: How They Impact Your Library, Presented by Jane Ryland

Barbara L. Scheid

In her presentation Jane Ryland reviewed the Coalition for Networked Information (CNI) and the National Research and Education Network (NREN), "two of the most exciting initiatives (she has) been involved with in her twenty-five years of working in the information technology and higher education fields." Ryland also introduced many of the challenges that must be met in order to make NREN a reality.

CNI was established by CAUSE, EDUCOM, and the Association for Research Libraries (ARL) in March 1990. The mission of CNI is to "promote the creation of and access to information resources in networked environments to enrich scholarship and to enhance intellectual productivity." Ryland reported, "CNI closely relates the mission of (our) libraries and higher education institutes which support and disseminate knowledge."

The National Research and Education Network is the product of the National High-Performance Computing Technology Act of 1991, which recently passed in both houses. Ryland said the intent of the legislation is to build on existing infrastructures including NSFNET, regional networks, and all the connectivity that is currently available. The network will include common carriers and private lines. NREN is meant to be a partnership among government, education, and industry.

According to Ryland, the future for higher education, libraries, and information and knowledge dissemination is the network: "The convergence of computers and communication technology will make this network possible." She cautioned librarians and information specialists to examine the current structure of the library within the information industry in order to maximize the benefits of the network. She said the library industry must successfully resolve issues, like the ones introduced at this conference, before it can consider how the network will help it.

Ryland stated that three of the major concerns to librarians are (1) there is too much information available, (2) information printed in published form is highly specialized, and (3) information is too expensive. She said 20 percent of the purchased materials are circulated 80 percent of the time. Eighty percent of the materials purchased are used by only two hundred to three hundred people in the world, and 60 percent of the materials purchased don't even circulate after the first few months they are acquired. "How can (we) afford to continue to buy and put in our libraries materials that are underutilized and so expensive?" she asked.

Ryland said at the current rate of increase of materials, the library will be the largest building on campus and will still be unable to house all the materials it owns. She said it is unrealistic to continue at this pace and that it is hoped the network will help. She indicated the network will help reduce costs associated with procurement and storage of materials by allowing a patron access to information that is not necessarily owned or stored by the library. NREN would allow a patron to browse the database and to select from the network only the information he or she needs. Ryland proposed that electronic storage would also help in preservation of information. She said, however, that the network is not a panacea, and there are many other factors that must be examined in order for NREN to be a successful solution to our present and future needs.

It is imperative to establish what the charges will be for information on the network. Ryland said it was the dream that CD-ROM technology would provide at a very low cost an unlimited personal library. The value of the information was not considered. Ryland noted: "The owners of the information as well as the value adders put a cost on information; and this is fair. (We) must be careful that (we) don't confuse the possible price reduction that

---

Jane Ryland is President, CAUSE.

new technology allows with the fact that there is a price to pay for having access to valuable information."

Ryland asked the audience to "imagine having been a professor of modern Eastern European history during the last couple of years without the availability of electronic information." Information needed to teach this course today is only available electronically. Much of the controversy over cold fusion was disseminated by electronic mail over Internet and by fax. Access to timely information such as this is important to the advancement of research and learning. It raises some basic questions such as how will this type of information be authenticated and archived and by whom, so that it will be available for future scholars?

Ryland stated that another issue of concern is how a user will find the information he or she needs. She said researching the floods of information available through the network "has been likened to trying to drink from a fire hose." A user's search on a particular subject may come back with one thousand or more entries. She said the Corporation for National Research Initiatives as well as individuals are working on a new technology called Knowbots, which should help users find the information they need. The Knowbot would help navigate electronic information and storage reservoirs. With this new technology, users would specify through the computer their particular interest, and the computer would survey the database and report what it found.

Ryland said it is evident that traffic over current networks is increasing every day and that there is a need for high-performance computers and a national network. However, too many people get caught up with what the network can offer. She said, "We need to moderate our use and choose when it is appropriate to apply the new technology and when to use other methods." Ryland argued that colleagues in the same office who communicate with one another by electronic mail are abusing technology. The author's intended form for an electronic newsletter is lost when colleagues print it, staple the thirty to forty pages of screen dumps together, and circulate it among the staff. Just as television did not replace radio and the VCR did not put movie theaters out of business, electronic information will not replace the book or human interaction. As Ryland put it, "If I thought that technological advances and global networks would mean the demise of books, I would stop working for it right away. I can't imagine curling up with a good workstation."

Some of the issues to be resolved before the network can become a reality are accessibility and ease of use and authenticity and affordability of the information. Ryland proposed that standards be created for the network regarding how information will be accessed, formatted, transmitted, and printed. Decisions must be made as to how to pay for the information and who must pay. If the cost of the information is too high, it will not matter how easy it is to access.

In Ryland's opinion, the National Research and Technology Network will enhance the competitiveness of the United States; provide access to super computing; remove constraints from research in terms of time, access, and distance; increase collaboration across as well as between professions; and enable industry to retrieve timely information. The capabilities of the network are numerous. Ryland optimistically stated, however, that "history shows that most of the uses of advanced technology are totally unpredictable ten years out so the availability of these capabilities will open incredible opportunities for us that we can't even imagine today." ■■

The Role of NCLIS and the Federal Government in the New Library Paradigm

Peter R. Young

One noteworthy feature of this meeting is the great variety of information professionals that are in attendance representing a wide array of interests and organizations. The event brings together a broad spectrum of library and information professionals from a diversity of types and sizes of information organizations and institutions. The richness of this mixture will, I am certain, make for a truly informative and enriching experience for all participants as we explore where we are and where we are going. Responsive leadership such as Dr. Chacra's vision to organize this conference assures that the professions of librarianship and information technology will develop together, as we must, in order to continue to serve the growing information needs of our rapidly changing society.

Truly, this broad representation of library and information service environments at this conference underscores a trend fostered by the increasing pervasiveness of technological change. Information technology has evolved over the past decade to the point that it serves as the basis for linking us together in communities composed of different specialists and disciplines. Linked digital networks are indeed going to dominate the 1990s. These networks blend different communities together and encourage a shared concern for shaping the information environment. Networks produce an enriched opportunity for extended dialogue. If we as librarians and information professionals are to survive and evolve in the emerging library and information service network paradigm, we must look for creative opportunities such as this meeting in order to foster collective interaction among the many different constituencies with concern for the future of our information institutions and policies. Our successful evolution involves the formation of alliances among many different types of organizations. Alliances that are broadly representative of all the different interests in our society. Occasions such as this conference afford opportunities to discuss and to dialogue about our collective futures and to explore visions of the new library paradigm. Our explorations will result in the actions necessary to build our future together as partners in meeting the full spectrum of information needs in our pluralistic society.

Despite the lovely setting here at Mountain Lake, my assignment this morning is not easy. In fact, it is among the most difficult public-speaking assignments it has been my pleasure to receive. When Vinod phoned me last May to discuss the conference, he suggested that I might want to talk about the role of the U.S. National Commission on Libraries and Information Science (NCLIS) in the future. I was excited to carry the NCLIS message to you, and I agreed to participate.

Later, I found out that what Vinod actually intended was that I address a much broader topic: the future role of the entire federal government in the area of library and information services. This assignment is not a simple task. It involves a complex array of topics covering intricate areas that only become understandable after a lifetime of involvement and commitment. It involves the federal government, both as a producer of information as well as a key player in the information policy-formation process. This broader topic is, in a very real way, a piece of the core mission of the NCLIS. After working for more than a decade in the federal government, I am more a student of the complex subject of federal information programs and policies than an expert in the relationship between the federal government and the nation's library and information services policies. After more than two decades, the NCLIS is well positioned to play a leadership role influencing the national policy that will affect libraries and other information service

Peter R. Young is Executive Director, U.S. National Commission on Libraries and Information Sciences.

organizations well into the next century. We are especially excited about two important developments this past summer. First, the successful White House Conference on Library and Information Services (WHCLIS) held in July in Washington. Second, the passage last month of some technical amendments to the NCLIS enabling statute.

My remarks this morning center on those federal programs and initiatives that will impact the developing library paradigm of the twenty-first century. I will give more detailed attention to the work of the commission and our plans for those specific federal programs that appear, at present, to have a distinct bearing on the emerging library paradigm of the future.

First, however, a little about NCLIS. The U.S. National Commission on Libraries and Information Science is a permanent, independent executive-branch federal agency composed of fifteen commissioners who are appointed by the president for five-year terms. The commission is supported by a small staff that I direct. In addition to the librarian of Congress, the law calls for five of the commissioners to be professional librarians or information specialists, with at least one of these knowledgeable with respect to the technological aspects of library and information sciences.

NCLIS has primary responsibility for developing and recommending overall (a term in the law that is very appropriate here) plans and policies related to library and information services adequate to meet the needs of the people of the United States. (If you ever needed a broader mission, I don't know what word you could put into it, but it's quite an interesting one.) NCLIS advises the president and Congress on the implementation of national policy, the need for cooperation among federal, state, and local governments, and public and private agencies, in assuring (here's another trigger word) optimum utilization of the nation's educational resources. Between those words *optimum* and *adequate* I don't think I have rested over the past year in terms of trying to figure out what the program should be, could be, and might be.

In fulfillment of the statutory mandate contained in Public Law 91345 (July 20, 1970) NCLIS, first, conducts studies, surveys, and analyses of the information needs of the nation; second, appraises the adequacies and deficiencies of current library and information resources and services; third, evaluates the effectiveness of current library and information service programs; fourth, develops overall plans for meeting national library and informational needs; fifth, advises federal, state, local, and private agencies regarding library and information services; sixth, promotes research and development activities; and, seventh, publishes reports on activities in these areas. This is an incredibly rich agenda, of which, over the past twenty years, I believe we have accomplished about one iota. In the next ten years, I hope we will amplify this package by a factor of one million. I would like to think that.

A technical amendment to Public Law 10295, which President Bush signed into law on August 14, makes five changes to the 1970 law, including the elimination of the ceiling on annual appropriations authorized for carrying out the work of the commission. By removing this cap on federal funding, and by adding language permitting the commission to receive in kind as well as monetary contributions, these technical but important changes place the commission as an agency in the position of being able to increase programming in the areas directly related to the merging library networking paradigm.

Over the last two decades, the national commission has initiated programs and projects in the following major topic areas:

- Exploration of the application of new technology to library and information services
- Work with the registrar of copyright to resolve intellectual property ownership issues and to revise the copyright law
- Examination of the needs for standard network protocols
- Oversight hearings on the Library Services and Construction Act and amendments to national library and information service legislation
- Studies performed by a task force on public/private sector relations

- Coordination of a panel on the information policy implications of archiving satellite data
- Cosponsorship of a symposium on information literacy and education for the twentieth century
- Development and adoption of the Principles of Public Information, identifying eight principles about the creation, use, and dissemination of information that is produced at public expense and owned by the people
- The conduct and planning of two White House conferences on library and information services, the first in 1979 and the second in July 1991
- The cooperative development of area statistics programs for recording, collecting, and publishing data allotted to public, academic, and special libraries and schools at the national level (this program is in conjunction with the Department of Education's National Center for Educational Statistics)

With broad statutory authority, NCLIS programs have encompassed a wide range of national policy issues and topics over the last twenty years. The commission has a unique policy-advisory role within the federal sector. NCLIS has the ability and responsibility to plan projects and programs related to developments affecting the evolution of what this conference terms the "new library paradigm." The form and shape of this new library paradigm constitute the central concern of this conference. Identifying attributes of this paradigm are also central to the program agenda of the national commission. Our program role in the topic of this meeting—"Linking Multimedia Digital Libraries"—is to provide the advice and planning required to assist the nation in achieving the vision of a nationally integrated network of libraries exchanging multimedia information in digital and other formats.

Before discussing the federal role in this new library paradigm, I want to try to draw the outline of what the new paradigm concept of a library actually looks like. I think there are few of us that doubt that a new paradigm is currently evolving for our information organizations and institutions, as well as for those of us who work in this field. What we are unsure of is what it feels like to work in the paradigm and what changes we need to initiate to achieve it. Symptomatic evidence of our current transitional environment is commonly accepted to be such items as the following:

- Increasing involvement of digital electronic systems in library and information organizations
- Movement toward networked information systems and a shift away from resource ownership toward access to shared resources from diverse collection sources
- Shift toward a more fluid format, where information is available in a multiple array of formats, both graphic and digital
- Shift toward greater user choice and control of the form and content of information and away from system constraints
- Rapid changes toward a global perspective on information issues
- Proliferation of information production that necessitates an increase in the ability to process and manage information at an ever-increasing level of efficiency
- Concern about the social, economic, and legal impacts of information illiteracy and electronic illiteracy
- Growing concern over the economic consequences and property-ownership implications of a transition from a graphic print-based information society to a culture dependent upon electronic communication as a means of information transfer and exchange
- Increasing concern about individual privacy and confidentiality, especially in relation to personal information and to issues involving individual information selection and choice
- Increasing involvement of different teams or groups of individuals engaged in collaborative work using information technology that bridges time and geographic distance, allowing cooperative group intellectual research and problem solving
- Growing awareness that the transition to a new library paradigm requires a careful examination of the inherent values underlying information creation and a shared exchange of the information processes

A paradigm is an outstandingly clear or typical example or archetype that establishes a pattern against which we measure change. We are addressing the new library paradigm at this conference, but, indeed, the paradigmatic shifts involved in the transitional decade of the 1990s may indicate a more fundamental transformational shift of power in society. Powerful new electronic tools and technologies are having a profound effect on global politics, social change, and our information institutions. Fundamental changes are occurring in the way we relate to one another. Individuals now have the tools to engage in collective group actions that tend to blur the traditional boundaries and identities we are so used to. Evidence of this is to be found in CNN's coverage of Operation Desert Storm in the Persian Gulf. Instantaneous transmission of the changing military situation allowed world leaders, along with cable subscribers, to see the effects of tactical decisions instantly. Further examples of fundamental communication changes are found in the reporting of dramatic changes occurring in the Eastern European nations, the dramatic changes occurring in the political structure of what was once the Soviet Union, and in the People's Republic of China during the spring of 1989, when Chinese students in the United States transmitted by facsimile Western news coverage of events in Tiananmen Square directly to students participating in political demonstrations in Beijing.

Technological changes are having an impact on the traditional functions and services offered by our libraries and information organizations, but these same technologies are also influencing our basic social structures in ways that are not clearly evident. Our paradigm shift toward the library of the year 2000 depends upon our ability to understand how basic values and principles are affected by the adoption of new technologies. These historic values and principles have served as the foundation of library service, and these same tenets must be translated into the new era. Perhaps the library information program is experiencing a shift toward what could be termed a "knowledge orientation" function rather than to the library of the future where more traditional services and resources are available on-site. Managing knowledge resources may, in the library of the future, involve new techniques and abilities that are very different from those acquired with our M.L.S.'s. We should be guided by visions of the future that define what we want to have available, rather than project future systems based on past organizational structures and functions. We should also welcome opportunities such as this conference to engage in dialogue about what values our future libraries should reflect. Only in such an open and interactive community can we develop visions of the future that accommodate the widely diverse, pluralistic, global society of the twenty-first century, where change is the only constant.

Symptomatic changes in the structure of the larger information marketplace can be seen from the perspective of the growth of the personal-computer marketplace over the decade of the 1980s. When IBM introduced its first personal computer in August 1981, the company hoped to sell 250,000 machines over the life of the product. Today, a decade later, this product has succeeded far beyond IBM's wildest expectations, with millions sold and with computers now accepted as a basic fact of life—almost as basic as electric power.

But, according to Paul Saffo, a researcher at the Institute for Future in Menlo Park, California, as noted in the *New York Times* ("The Computer Age: Still a Work in Progress," Aug. 11, 1991, Section 4, p.1,4), "We're mostly just doing old things in a new way. The main event is yet to come." Indeed, just ten years ago, before personal computers were introduced into the U.S. market, there were just 230,000 computers in use in this country. Today there are more than 55 million in our country and more than 110 million in the world.

So where are we to find evidence of the new paradigms for electronic libraries, digital business systems, paperless offices, and an educational system transformed by computer-assisted instruction? What miraculous, clear new archetypes will result from the digital information-processing revolution of the last several decades? Perhaps, in looking for the *big* changes, we are

missing important clues to the future impact of electronic technology on our information systems. For example, claims of improved productivity through the introduction of computers have not, in general, proven accurate. Service-industry productivity in the last decade has risen less than 1 percent a year, compared with 4 percent in manufacturing. Part of this low increase in productivity despite the investment of hundreds of billions of dollars in computers may be that we are spending more of our time making revisions and fine-tuning our information, simply because our information tools make such refining easy to do.

In a very real way, the claims of revolutionary development and the emergence of new information paradigms that were made at the start of the last decade, will, indeed, become evident in the next ten years. The interconnected national network that has been described as the "electronic superhighway" will, in effect, provide an infrastructure for exchanging messages, engaging in collaborative work, customizing human interaction, and extending the capability of machines. With computers hooked into national digital networks, the isolation of intelligent machines will be broken, giving way to new breeds of intentional interactive communities. At the same time, as digital highways link individuals and institutions, a fuller spectrum of format communications becomes possible, allowing voice, data, and images to enrich the complex texture of human interaction.

At the same time as our postindustrial society enters the age of networks, linking intelligent human agents into global communities, our social, economic, and technological policies will require attention and revision in order to assure that our values and principles are consistently carried forward in the transition to a network literate society.

The visions of the library and information service organizations of the future can be seen from the standpoint of three current NCLIS concerns: (1) the White House Conference on Library and Information Services, (2) the emerging national initiatives relating to digital electronic-network technology, and (3) the changing global economic climates affecting the legal, technical, and institutional structures of our information organizations.

The nation that received the 1979 conference resolutions from the first WHCLIS is very different from the America we know today. Our roots and values are not different, but our public policy concerns have shifted. Three concerns that had a profound affect on the 1991 conference recommendations are (1) America 2000, (2) the proposed National Research and Education Network (NREN), and (3) the increasingly global nature of information issues.

AMERICA 2000

In 1991, one of our most important current national priorities involved a long-term strategy to reinvent the American educational system—to design new American schools for the year 2000 and beyond. This challenge to revolutionize American education involves a partnership among national, state, and local officials, along with the corporate community. The strategies and goals involve national standards for American schools and nationwide assessments for measuring student progress. There are implications here for libraries in words such as *assessment*, *measurement*, and *evaluation*.

The education strategy expressed in America 2000 involves a change in our attitudes about learning. This is not a federal program. It involves a populist crusade to spur fundamental changes in the way we educate ourselves and our children. The lesson for those of us concerned about improving the nation's library and information services is that our conference implementation strategy for the recommendations of the White House conference must involve the broadest possible range of interests. To be successful, we must use the nation's educational revitalization efforts to leverage the conference outcomes for improving our libraries and information services that are essential to the success of programs.

NREN

Our public discussions about library and information service issues in 1991 were dominated by what has been called the

high-speed, high-capacity national digital interstate network, often compared to the interstate highway system, not for the transportation of goods, but for information and communication services involving different types of media. Here again, the strategies for NREN implementation involve discussions identifying the various complementary roles of the public and private sectors in developing, managing, and maintaining this new national communication infrastructure. Once again, the lesson for those of us concerned about improving the nation's library and information services is that our conference and implementation conferences such as this and recommendations coming out of the White House. Strategies for ideas and actions coming out of this event as well as the White House conference involve the broadest possible range of interests, interests representative of the broad array of individuals attending this conference. To be successful, we must use the national network planning efforts to leverage the conference outcomes for improving our libraries and information services.

THE INCREASINGLY GLOBAL NATURE OF INFORMATION ISSUES

The nation's focus on creating a revolution in education is not concerned only with student success and achievement. It is motivated by a concern about which nations will thrive in the competitive global marketplace of the future. Increasingly, in the changing global information structure, dominated by technological change, knowledge and ideas are seen as the raw material for growth and economic achievement. Nations that nurture ideas and the means to express knowledge will move forward in years to come.

FEDERAL FUNDING OF LIBRARIES

Finally, I would like to share a moment's reflection about the national climate and fiscal reality. I had lunch last month with an extremely knowledgeable person who works as a budget examiner at the Office of Management and Budget, the infamous OMB. During our conversation about the relationship of libraries to the president's education agenda, she made a few thoughtful observations about the next few years that included the following:

1. We are in an era of reduced federal spending.

2. This era is made more complex by the budget reform that divides the federal budget into four distinct parts related to domestic, foreign, defense related, and entitlement programs. Increases in federal support to a program in one area must be taken from other programs within that same quarter.

3. Program growth in the next few years will lag behind inflation.

These budget realities make it extremely difficult to initiate new programs or to plan for increased support for existing programs. The U.S. economy depends on rapid adoption of new information and technology. Over the past two decades, the dollars spent on federal programs for education, training, employment, and related social services fell by 21 percent when inflation is taken into account. This is the climatic context within which the one hundred White House conference recommendations will have to be considered as well as the context within which the new library paradigm must be viewed. Fiscal restraint dominates our planning due to the $360 billion national deficit. Until relief is gained through new revenue sources and structures through which a mixture of private and public support for libraries and information services is achieved, we will not likely find support for implementation actions geared toward the development of new library network paradigms at the federal, state, or local levels. This is why the library community must form strategic alliances, such as with the coalition of allies in the information services sector.

John W. Gardner, founder of the lobbying group Common Cause and professor of public service at Stanford University, addressing the one-hundredth commencement of Stanford this past June, talked about the nation and the test of character we face:

> Life isn't a mountain that has a summit. Nor a game that has a final score. Life is an endless unfolding and, if we wish it to be, an endless process of self-discovery, an endless and unpredictable dialogue between our own potentialities and the

life situations in which we find ourselves.

The test is whether in all the confusion and clash of interest, all the distracting conflicts and cross purposes, all the temptations to self-indulgence and self-exoneration, we have the strength of purpose, the guts, the conviction, the spiritual staying power to build a future worthy of our past. ■■

Questions and Answers with Jane Ryland and Peter Young

Question: We have heard that the National Research and Education Network (NREN) started out to be the core of high-performance computing and that by putting the *E* in NREN it's been expanded to include education. How do both of you see it expanded to all types and sizes of libraries and to individuals in their homes? Is there any thought of that next stage?

Jane Ryland: There are many of us who believe NREN won't be effective until it reaches libraries of all types and sizes and individuals in their homes. Libraries increasingly are not buildings with walls. Libraries are points where you can access information resources. You should also be able to access these resources from your home, just as you can from within a physical structure. However, there are varying degrees of acceptance in this belief. The question is not if there will be equal access to the NREN, but how quickly it will evolve. Clearly, institutions with the most resources will be the first ones to get access to the capabilities, but I think that there is a strong belief that everyone will eventually have access to the information.

Peter Young: I attended a preconference to the White House conference in December that was managed and organized by the Network Advisory Committee of the Library of Congress. The conference was attended by 125 people and addressed the future role of networks. I was amazed because in two and a half days that group, composed of primarily academic and network leaders, endorsed a concept that public libraries need to be nodes on a national network. That endorsement was fed into the White House conference and produced the second most popular recommendation coming out of the conference.

I will read it to you. It is entitled "Network Recommendations":

> Congress shall enact legislation creating and funding the national research and education network that will serve as an information super highway and allow educational institutions including libraries to capitalize on the advantages of technology for resource sharing and the creation and exchange of information. The network shall be available in all libraries and other information repositories at all levels. The governance structure for NREN shall include representation from all interested constituencies including technical, user and information provider components, as well as government and education at all levels.

That is a dream. How to take that dream from a political reality to a social reality is the challenge of this community.

Question: We have learned about both ends of the NREN issue that seem to be at opposite sides. On one hand we are talking about making information over the network available to everyone, and, on the other hand, we are talking about adequately compensating the people that generate the information. I am wondering how these two opposite sides of the coin are going to work together. I think one specific instance is the CD-ROM products available today that were marketed with the idea of being used at one location or, at the most, on a local area network. These databases are now going to be accessible over an infinite area to an infinite number of users. How are the people that generated the information in the first place going to be adequately compensated?

Jane Ryland: There is no easy answer, but the Coalition for Networked Information (CNI) realizes that this is an extremely difficult question, and groups within the coalition are working on it. Copyright legislation is based on printed products. But

what we have now is the capability for incredibly broad distribution of information via the network and distribution without anyone's knowledge. The (copyright) laws are going to have to change. A little bit of background may help in providing a base for discussion. Prior to the 1950s, 80 percent of the revenue for information went to nonprofit organizations and professional societies, and 20 percent went to commercial publishers. Now that ratio is reversed, and it's not clear whether all the compensation in today's system is going to the authors as opposed to the (publishers and printers). Sometimes authors pay for the publication of their works rather than get compensated for them.

Peter Young: This question is a particularly interesting area of the NREN, especially when you go back not ten but twenty years. Look at the relationship of the Association of American Publishers (AAP) the trade publishers as well as the scholarly publishers, and the library community. It was a very different world in the late 1950s up to about the early 1970s. About 1978 things started to move apart, and since then, technology has put up additional barriers or introduced walls between the two communities. Part of that I think is the result of the publishers' being in a more conservative environment than even libraries. That conservative environment sees technical training not as an avenue for amplified products and services with increased revenue sources and strings but as a threat to revenue sources and strings. This is a very significant threat. How can you control ownership of a single work that isn't expressed in a tangible medium? What will happen when, by a very simple action, an individual can be an author, a publisher, and a distributor and/or a consumer of a revised work on a network? There are a variety of different models you can look at for exchange of dollars via the following on the exchange of information products on a network. Many of those models come out of the European environment in which you have a third-party-right balancing organization like the Copyright Clearance Center and some of its experiments. This solution introduces a certain amount of overhead and the technological necessity for putting some sort of meter within the network environment. A meter would track use, presumably collect sources like a performing-rights organization, and dispense royalties to the owners. As Jane pointed out, the owners of the information are currently the publishers, rather than the authors. What technology has done is give a lot of power to the individual to supplant those institutional functions that have been fulfilled in the past by publishers and libraries. The main question is: What is the role of the publisher in the future? I think Jane put her finger on it when she started talking about authenticity. There are very few professors in academic institutions who will turn to the network for tenure publication consideration without going to an Elsevier or Pergamon, especially in the scientific, technical, and medical fields. These concerns are policy related, and technology is in front of the policy changes. Right now, I am very hard-pressed to say if the world were mine tomorrow, what changes I would make to the copyright law.

Jane Ryland: Part of what the coalition is doing is developing vision statements for different kinds of models. A very interesting vision presented at a recent coalition meeting described the national licensing plan where publishers or providers of information would give unlimited use of materials to the subscriber who would pay a couple of dollars a month for the use of all types of materials. There are many alternatives that might valuably be discussed, and ultimately the determination of how much compensation should be paid will be made by a very diverse group that represents everyone involved. A coalition project through Elsevier and a group of institutions is pioneering a publishing project. Elsevier is working to make a critical mass of serials and periodicals of several disciplines available for a fee. The institutions or libraries buying this particular service will pay the same price whether it is a printed publication or whether it is an electronic publication for which there will be unlimited use of that particular material.

Question: It is only fair to make the information on the network available only to those who have the ability to pay. However, this philosophy will inhibit the desire

to make the network accessible to everyone. Therefore, will the entrance routes from counties to the national highway be federally supported in order to provide equal access to the network or locally supported?

Peter Young: I came down here from Washington the other day and was overwhelmed by truck traffic on Route 81. I had planned about a five-and-a-half-hour trip, but ended up taking eight hours. The reason we took eight hours over the five and one-half is that 81 is a very heavily traveled interstate, especially by trucks. The reason 81 is so busy is because the commercial traffic is shying away from 95 because the tolls are being raised. We also drove to Vermont last week. We shunned 95 along with all the trucks because the New Jersey turnpike just tripled its rates. The point is that traffic along a locally managed, pay-as-you-go, fee-based highway is such that it's driving users to publicly subsidized alternatives. What will happen if all of the digital network gateways are controlled by private-sector interests? Look at what happened to commercial television. I have been trying to write a piece of a model that describes what conceivably could take place in an environment in which you have a Corporation for Public Broadcasting type of model for scholarly communication within a digital environment. It is very hard to see how a mix of public, federal, private, and institutional not-for-profit representative interests work together for finding their way in this dark forest. The questions I have been asking myself are: What are you going to do about ownership? How do you compensate authors? How do you make sure that the balance between incentive for individual profit-making entities is satisfied and at the same time the public interests are served? These are very hard, but extremely fascinating questions.

Question: CAUSE and EDUCOM are pretty representative of their user community and are good organizations. However, ARL is certainly not representative of the interests of most librarians. It is, therefore, highly questionable that representatives from ARL are the ones calling the shots on the NREN. How can we sit here today and talk about how the development of the NREN is represented by diverse interests when there is little representation from smaller organizations and libraries?

Jane Ryland: CAUSE is probably the most broadly representative of the three organizations you stated, with a larger component of smaller institutions and community colleges than the other two. However, the members of all of the organizations feel it is important to get all the various players involved in the development of the NREN and are using various mechanisms to do so. The Coalition for Networked Information was formed especially to address the needs of all types of libraries and information agencies. Even though the dues to become a member of the coalition may be a barrier, an institution may still be a part of the working groups of the coalition. These working groups are where much of the actual development of policies occurs. The coalition has also encouraged a number of small institutions and small libraries to form consortia (sharing the dues) so that they could participate. The publication and communication mechanisms of the organizations are also being used to try to keep the broad community, not just members, informed about the issues. The CAUSE conference has invited a constituent group to its conference no matter what their involvement in the coalition to give some feedback about the NREN to key coalition members. ARL, even though it only has 190 members, is also trying to get word to all libraries through its publications.

Question: How do you see technology modifying the library's objective of efficient widespread dissemination of information, and how will it modify the way in which the library operates? Also, do you think the librarians' role is going from a traditional institutional function to a teacher role?

Peter Young: From a certain perspective librarians may think we have had our warehousing function somewhat stolen. When Vinod was talking about the comparison of where things were stored, in terms of the storage-space capabilities of video disks, etc., I kept waiting to hear about libraries as a means of storing information and the millionfold factor that libraries give to information from the past and to images, thoughts, and ideas stored

from those who have come before. How can we evolve these institutions into things that are indeed efficient and effective at our traditional roles as those roles become encroached by other things in society and as information plays a more public role in societies? If I had to identify a niche market for the librarian, I would identify the classroom.

Jane Ryland: I don't think technology and technological capabilities are going to be something that makes us change. I think the librarians will decide what the best uses of technology are and how their own roles will change. Given the capabilities of technology, I think the librarian will have a choice to make and will have the opportunity to determine the best way to utilize the technology and decide how the library as an institution will change within society. ■■

Library Automation Marketplace: The New Frontier with Fuzzy Edges, Presented by Frank R. Bridge

Barbara L. Scheid

Frank Bridge's presentation gave a review of the library industry ten years ago and how it is evolving today and provided a glimpse of what will happen in the future. He also gave suggestions on how to plan for library automation.

Bridge said librarians have been refining the information storage and retrieval process for more than three thousand years. Librarians receive the most out of new technology because they have already implemented standards. He said, "Librarians are the most informed people about how they want their information systems to work and what standards the new technology must meet. Librarians should not rely on information technologists, but should plan their own futures."

The audience was briefed by Bridge on how information technology evolved and changed the way in which patrons use the library. He said connectivity has been and will always be an automation concern. For example, first a system consisted of only a central computer with hard wires connecting dumb terminals. He said it was easy to determine where the system began and ended and how many users were on the system and where those users were located. Very quickly, however, hardware and communication systems were improved so that users could dial into modems attached to the computer. A patron could now search the system without entering the library. Bridge stated, "At this point the library ceased to be a building and began being a service." Local area networks within the library were then added to the system, and before long, analog lines and some fiber optics were laid that connected off-site LANs and other libraries to the central system. Through the use of digital lines, the central computer was then attached to computers in other cities, states, and countries. A patron of a library located anywhere may now search a library catalog and the external databases the library subscribes to without leaving the library-automation system. Bridge said, "Presently, it is getting very fuzzy to determine where the library system ends and other systems begin."

Bridge stated that although the library-automation business is changing rapidly, there are things a library can do now to plan for the future. He said, however, "It is not possible to have a ten-year automation plan. Librarians need to plan much more closely to the future. Day-to-day operations and the services a library offers are changed by technology too frequently to plan ten years into the future."

Libraries that have not yet automated or

Frank R. Bridge is President, Frank R. Bridge Consulting, Inc.

libraries wishing to automate operations further are faced with questions as to which library automation software and hardware vendor to choose; how the software vendor handles electronic mail and multimedia; and how both the hardware and software company handle connectivity issues.

Bridge said the library should research communication devices to see which are best suited for it. The library should consider how many branches the system will support and whether the central system will be shared by other libraries. He added that the library may consider using locally mounted databases and should know how their system will interface with these databases and with other systems.

According to Bridge, "Hardware-independent operating systems are a myth right now. The UNIX operating system is really not hardware independent. It must be fine-tuned for each system." He said that in the future automation companies will handle the hardware-independent operating-systems issue. Hardware companies such as Hewlett-Packard and DEC are rewriting their operating systems to meet the standards for UNIX. He also asserted that "the standards for UNIX are going to be developed all over again." He said, "There is a development path to migrate systems like MPE/XL using the Posix guidelines, but I don't know the development time frame."

Bridge recommended purchasing either an Ethernet or Token Ring LAN, as most systems are compatible with these two LANs. He stated that the most important decision when planning a LAN is the type of connection to purchase for connection to the cabling. He stated that twisted-pair cabling is cheaper and gives some flexibility. However, in order to keep it separate from the phone lines one needs to make sure that the LAN twisted pairs are a different color from the phone-system cabling.

Librarians should also consider how their vendor is planning for electronic mail. Bridge said this is a concern from the point of the local application software, to the computer system, up to the connection to other computer systems. He noted that X.400 Electronic Mail is coming along very rapidly and works well within systems. He said most vendors are writing software to make it work between systems, and this should be working quite well within the next year or so.

As Bridge stated, the Fiber Optics Digital Data Interface (FDDI) is the standard to move data across the networks. The bandwidth on fiber optics is so great and can carry so much data that the investment a library may make today will be good for a number of years. He added that telecommunication companies are now offering "bursty telecommunications." Bursty telecommunications allow a library automatically to gather up just the right amount of bandwidth when it needs it and then decrease the bandwidth to normal levels after transmission of the files. This service is particularly useful for transferring large files.

Bridge asserted that although vendors give high praise to Open Systems Interconnect (OSI), they really haven't done much work with it and "file server location information" software is much more popular with the vendors. He reported that many library automation vendors are now writing software that includes in the subject headings item records for images and other multimedia collections that link the item back to the subject headings. There is no need to get out of one system and into another. Bridge said this new technique is a major development in dealing with the OSI concept because the patron is going from one system to another very easily and quickly.

Some helpful hints were given for those libraries that have not automated and for those that are considering further automation. Bridge said the library system the library purchased should handle unabridged MARC records and should not throw out any part of the record. He recommended the use of Code 39, or CODEABAR, bar-code labels, as they are the most supported labels. He told librarians "not to use dot matrix printed bar codes, but to use laser-generated or photo-composed bar codes." He also said when libraries are sharing a system, it is very important that each library has a local library identifier on the bar codes or at least that different bar-code runs be designated for each library.

Bridge also suggested that libraries implement USMARC and LC cataloging practices whenever possible and keep stan-

dard bibliographic identifiers. He said that it is imperative to document every decision for cataloging and coding and that libraries should have written procedures for all automation procedures. A good example is to document how a library encodes information, particularly item-specific information. He also said it is important to catalog serials holdings at level 2 or above of the new MARC standard for coded holdings and locations for serials.

Bridge stated that technology has changed the way in which a library provides services to its patrons. Today's patron expects easy and prompt access to all types of information resources. Therefore, the librarian of today must be even more informed than the librarian of the past, because he or she must still possess a wealth of information and must also understand and know how to apply technology to meet the new demands of the users. ■■

The Vision of the National High-Performance Computing Technology Act of 1991

Rick Boucher

Increasingly, access to information and the rapid and innovative exploitation of information is integral to the creation of new wealth. Postindustrial societies have entered the information age. Advances in information technologies—computers, fiber-optic networks, high-definition video systems—are transforming both work and leisure environments and the way society operates. Some experts have characterized the developed world as experiencing a transforming convergence of computing and communications technologies that will have an impact as great as the replacement of muscle power by machines.

The government has always played a central role in the development and maintenance of the physical infrastructure of the nation—the roads, bridges, canals, railroads, seaports, and airports. In the information age, to ensure competitiveness in a global economy, the government has a legitimate role to assist in the development of the nation's information infrastructure, including digital networks, advanced computers, and distributed databases of specialized information.

Libraries constitute an essential part of the information infrastructure and will continue to play a key role in the advance of civilization by perpetuating and preserving knowledge. However, as suggested by the theme of this conference, the way libraries fulfill their function will evolve as information technology changes. As the sheer volume of information expands at a dizzying rate, the development and availability of tools to help users find and understand the information they seek will become more and more important. This is clearly a central service libraries will be called upon to provide.

The magnitude of the information explosion in science and technology has driven, and has been driven by, the development of powerful computers and high-speed data-communication links. Computers can create elaborate mathematical models of the world around us that allow detailed study of how the elements of the models interact in space and time, outside of the limits imposed by the real world. That is, computers can shrink oceans, zoom in on molecules, slow down physics, and fast-forward climates. The ability to see physical phenomena at the right size and at the right speed leads to understanding.

However, the immediate product of large computer simulations is a mind-numbing barrage of numbers. Similarly, experimental science has developed techniques and instruments that produce enor-

Rick Boucher is Representative, United States Congress, Ninth District, Virginia.

mous data streams. For example, a single remote-sensing satellite of the proposed NASA Mission to Planet Earth will produce, in five days, data equivalent to the entire contents of the Library of Congress. Likewise, the Human Genome Initiative, which has the goal of cataloging all of the genetic information that defines the human species, will establish an enormous database. This digital database will be the raw material researchers will access and manipulate in order to identify gene expression and gain understanding of genetic-based diseases.

Only by means of high-performance computers can these torrents of numbers be translated into visual images that can then, in turn, be interpreted and understood. However, to transmit large storehouses of data to the computers that convert the raw data into useful information and also to provide talented researchers with access to that information for analysis require high-capacity data links.

A data superhighway linking all regions of the nation would allow scholars from across the country to interact and collaborate as if across the hall from one another. It would spur collaboration among talented individuals irrespective of their geographic location and institutional affiliations, as well as allow easy access to remote databases and large facilities. Moreover, an accelerated version of traditional scholarship will be facilitated as noncomputer specialists are assisted in finding and acquiring information from distant locations.

Applications of a national network will eventually extend far beyond research and university-based scholarship. Schoolchildren will have new worlds of information opened to them to explore at will; physicians will be able to consult with specialists thousands of miles away to assist in interpreting a patient's CAT scan in the middle of an emergency; home-entertainment centers will include interactive digital television that lets viewers explore artificial worlds existing only in a computer's electronic circuitry.

The most exciting uses of high-performance computers and networks have not yet been thought of, however. This has always been true of revolutionary technologies. Consider that railroads were envisioned as feeder transportation systems for canals and that even Thomas Edison, a visionary of stature, thought that the principal use of the phonograph would be to record the final wishes of old men on their deathbeds.

The full promise of high-performance computing and networking is far from being realized. The current version of the national network, the Internet, is neither fast enough for computational scientists and engineers nor transparent enough for persons with limited knowledge of computers.

Applications may speed along one link of the Internet and slow to a snail's pace on another. Subnets constituting the Internet operate with different systems hardware and often with their own peculiar rules, or protocols, for formatting, addressing, and transmitting data. Not only do the complexities associated with communication on the current Internet take time from research, they also discourage many researchers from even using electronic mail.

Although much research and development work has yet to be done, the basic technological component needed to turn the tables on slow data-transmission rates is now at hand, namely optical fibers. The raw, inherent data capacity, or bandwidth, of a single glass fiber is a million times greater than that of a copper wire. Best of all, the intercity optical-fiber links have largely been put in place by the telephone companies, although the capacity available has hardly been tapped. The problem is at the ends of the fiber, at switching points and entry/exit ports, where electronic signals must be rapidly transformed to coded light pulses—and vice versa.

Research is now underway to increase the useful bandwidth available for networking applications with optical fibers, including early development of all-optical computers and switches. Current projections expect data rates of billions of bits per second over switched networks within the next five years.

As many of you are aware, congressional action has been taken this year in order to ensure that data superhighways become a reality and that the power of computa-

tional science and engineering continues its advance. The National High-Performance Computing Technology Act of 1991 was passed by the House with bipartisan support in July, and a companion measure in the Senate was recently approved. I believe that a compromise between the two versions of the bill will soon be achieved.

This legislation creates a balanced program with several connected components. It will develop the National Research and Education Network, or the NREN; it will explore new classes of high-performance computers, including development of software to exploit the capabilities of the most powerful machines for science and engineering applications; and it will train new computer scientists and computational scientists and engineers. The several components of the high-performance computing program will consolidate and focus federal research and development activities. Overall, the program seeks to capitalize on our lead over our international competitors in some areas and accelerate developments in other areas in which the competition is tighter.

An important focus of the high-performance computing program is to address particularly challenging problems in science and engineering that are barriers to scientific and technical progress. Those problems are often referred to as "grand challenges," and many are of great importance to society. Examples include development of computer simulations of climate to assess the consequences of human activities; analysis of the fundamental structure of materials to develop better high-temperature superconductors; and determination of the function of biologically important molecules to unlock the secrets of cell biology and, thereby, open new avenues for the cure of disease.

The National High-Performance Computing Technology Act has attracted broad support from industry and academia and from many types of users of computers and networks, from the occasional to the most sophisticated users. I am pleased to say that librarians, both individually and through their professional organizations, have been active in their advocacy for the legislation and offered many helpful suggestions during the process of developing provisions related to the network and to services available through the network.

The National High-Performance Computing Technology Act provides a framework for establishment of the NREN but leaves many of the details for implementation to be worked out by the agencies involved. Policies for user access, cost recovery, and security, to name a few, will need to be determined as NREN is established.

A critical aspect of this process will be the interaction among the cooperating federal agencies, the telecommunications industry, and the outside user community. The House included a provision in the legislation to create a formal advisory panel composed of nonfederal members to assist federal officials with planning and implementing all aspects of the program. The membership of the advisory panel is specified to include representatives of the research, education, and library communities, along with network providers and industry. The goal of this provision is to ensure that communication occurs among all of the interested parties in the public and private sectors. Good communication will help ensure that the design and implementation of the NREN results in the prototype of a truly national network, in terms of applications as well as connectivity.

The opportunities offered by the high-performance computing program are great, and we in Congress intend to move aggressively to implement it. The program is at the heart of efforts to enhance the nation's economic competitiveness and will provide resources to stimulate the creative imaginations of scientists, engineers, and educators, and all those at the forefront of the information revolution. ■■

Electronic Libraries: The Pros and Cons of Multimedia Access Copyright Issues, Moderated by Jane Ryland

Barbara L. Scheid

Jane Ryland opened the session by explaining that copyright issues apply to multimedia information as well as information of all kinds.

Ryland began the discussions by covering the background of the copyright law, and the topic quickly proceeded to copyright ownership issues. The audience also examined the advantages and disadvantages of current copyright laws and addressed the problems associated with copyright on electronic information.

The entire concept of copyright stems from the creation of the printing press, which actually made making copies much easier and made commercial copies widely available. The United States copyright law stems from the British copyright law enacted during the period of Queen Anne. The law was first seen in the U.S. Constitution and "secured to authors the exclusive rights to their writing." The current copyright law protects literary, musical, and dramatic works as well as pantomime, choreography, pictures, graphs, sculptures, motion pictures, audiovisual material, and sound. In 1976 the law was amended to include computer programs, electronic music, and computer databases. Copyright law includes the physical copying of material, but does not cover the use of ideas.

The intent of the Constitution was to encourage the creation of intellectual work. Scholars were to be compensated for their work so that more scholarly works would be created, but is this what is happening today? Authors do have exclusive rights, but there are exceptions to the laws, such as the fair-use provision and the exceptions made for libraries. It was stated that today most authors do not own their works; the publishers own the copyright to the works. Approximately 80 percent of the return of copyright revenues goes to the publisher. Part of the copyright intent was to reward the author of the original work and then the value adder. Is this added value worth 80 percent of the total revenues, as opposed to the value of the author's work?

Why, then, do authors give the copyright to the publisher? Does the writer still need a publisher? In the past the publisher took the risk and added value—like authentication of the text and creation of the market for the work. However, universities are the source of much of the authorship of intellectual work, and technologies are now available that give a university the capabilities of becoming a publisher. The peer reviewers of intellectual works are faculty members who already work for the university. In essence, peer review is the mechanism publishers use for authenticating the work. What is left for other publishers to do? It was decided that this point is bringing the publishers to the table. There are some publishers who know this is a problem and are working with the industry to ensure that they will have a future role, but there are many publishers who need to "jump on the band wagon."

Ryland reported that the Coalition for Networked Information (CNI) is providing at least one forum where publishers and university representatives will get together to communicate each side of the issue. With hope, creative solutions will be established by these types of forums.

Advantages and disadvantages of the copyright law were identified by the participants. The advantages mentioned are that the law encourages creativity; it rewards the writers who publish the most-used materials; it protects against unfair use of materials; and it helps the system of faithful reproduction of works. Some of the disadvantages of the copyright law are that the cost of administering the law is high; the language of the law is vague and perplexing to the average user of information; the law inhibits the free flow of information; and some public relations problems occur for the agency responsible for ensuring compliance with the law.

What does "fair use" allow a user to do with materials? A summary of the conclusions drawn from the discussions of the participants follows.

Under the fair use of the copyright law, a library may transfer information from one medium to another in order to facilitate its patrons' video and stereo equipment. Multiple copies of printed text may also be made, but not for profit. A professor may take small portions of a commercial video to teach a script class, but may not show the entire video unless the video is bought with public-performance rights. A taped television series may only be kept on reserve for one month. After that time, the library must buy the video from the producer or retape it when the series reruns. Copyright of art collections can also be very dubious. A library wishing to create an image database of artwork for art historians to search must receive permission from the owner of the artwork in order to have it in the database, as the database may be considered public domain.

The fair-use exception to the law has limitations. Journal articles placed on reserve for a class may not be reserved for two consecutive semesters unless the faculty member has written permission from the publisher for permanent reserve of the material. Making multiple copies of the printed text must not impede the market for the article, and the copying of the material must meet the test of spontaneity. It was stated that as a result of a lawsuit, Kinko's is setting up a copyright clearinghouse for faculty. Kinko's is trying to get as many publishers as possible to agree to the program. If successful, the program will make it much easier for professors to offer all types of articles to their students. Under the program, Kinko's is responsible for collecting a royalty fee from either the students or the faculty and for paying the royalties to the owner of the work.

It was noted that government publications are not covered under the copyright laws and may be copied. Information that is brought into a congressional hearing as a public document is no longer private, but public domain, and may also be copied.

It was clear from the discussion that fair use might be called upon to cover a multitude of situations. Guidelines for fair use do not explicitly cover the educational use of materials. For more information on the fair-use issue, Ryland recommended a book entitled *Questions and Answers on Copyright for the Campus Community*, written by the National Association of College Stores and the American Association of Publishers.

The fair-use exception in the copyright law may not always be applied and information is not free. How can a fee be collected easily to reward the author and publisher? The following models were given by the participants.

There is much support for U.S. legislation that adds a tax to the price of VHS tapes. The tax is distributed by the government as royalty payments. It is believed that this tax will help pay for the pirating of movies and other types of programs. Another method of collecting royalties is a device called the "demonstration package." This combination software/hardware package automatically accounts for copies made of magazine articles or of entire magazines.

Accounting for the usage of databases mounted on CD-ROMs and available through local area networks is a present concern. One of the pricing schemes envisioned uses a meter to monitor the use of various CD-ROM–based products. Under this pricing scheme a university or college library is charged for a range of users a month. The library then pays the difference for usage over the range and is credited for usage under the range. A software package located on the file server monitors the usage of the database. Unfortunately, this model will not work for databases loaded on the mainframe.

Vendors are having a very difficult time pricing the availability of databases on the mainframe. One vendor currently charges libraries by the number of students enrolled at the university. The argument against this pricing structure is that the whole student body is not interested in one particular database. It was proposed that the vendors price a specific database based on the number of potential users. Usage of the database then is monitored by a metering device attached to specific ports accessing the database. When users dial into the database, they identify themselves and the metering device starts. Of course, using this method, the library pays nothing for the item and

only for the usage of the item. The author and publisher receive their royalties based on the usage of the work. For example, if one thousand users request an author's publication, then the author will get the fixed, per-copy fee times one thousand. If no one requests the publication, the author and/or publisher get nothing.

Another vendor bases its pricing on the number of terminals that have access to the databases. This can cost thousands of dollars depending on the size of the library. These vendors are discovering that libraries cannot pay subscription rates based on the number of terminals. The new solution is to charge a fee based on the number of simultaneous users. This is much more reasonable, as a library may have two thousand terminals but only twenty-five simultaneous users.

It was suggested that the users pay for the use of these databases. Some participants were very concerned that the library industry would be so willing to charge for information when in the past it has not. It was thought that this method would also impede the use of the library and the databases. "In the past, a student's tuition covered unlimited use of the library's holdings. Now, the library will have to tell students and faculty that there is an additional charge for doing their homework and/or research," a participant said. The image of the library has always been that it was "free to the end-user." A participant continued, "Do (we) as librarians really want to change this image? There will be several public relations problems if the library becomes a fee collector."

The administration of copyright for electronic information is a major concern. National and global networks make this type of information available to higher-education institutions, libraries, and businesses around the world. The National Research and Education Network will, in the future, make information available to people in their homes.

Regional models are being developed that will help to handle the authenticity and flow of information. Regional centers, each representing thousands of possible users, will be created. These centers will collect fees and distribute license fees to the owners of the information. Ryland said, "It is hoped that by collecting a very nominal annual fee from users of the network, the cost of the information will be spread widely. Therefore, no one will be burdened by the cost, and the owners of the information will still receive the same amount of money."

A very important aspect of electronic information is the authenticity of the work. A user of the network will have the ability to gather segments of authoritative work on a particular topic, add a little value to the work, and then publish the work as an original. Some technologies have been developed to restrict a user from doing this, but there are none that a "hacker" can't get around. It was also mentioned that standards and policies must be developed for archiving this information. "It is almost getting to the point that the state of the art will infringe on the state of academic research and publication," said a participant.

There were varying opinions on what the librarian's role should be in developing policy and copyright laws. Some stated the librarian's role should be to educate and promote the standards and laws developed by groups like the CNI, CAUSE, and EDUCOM. This group thought that participation in the development of the policy and copyright laws would only encourage policymakers to make librarians the "copyright cops." Many others in the audience disagreed with this approach and thought that librarians should first be educated about copyright law and should then be one of the main players in implementing changes to it. It was suggested that library schools offer a class on copyright law.

Ryland concluded the session by explaining that at least 80 percent of a library's user base must have electronic access to the information on the network in order for it to be economically feasible for the library to offer the information electronically. She said, "Until eighty percent of the user base has electronic access to the information, the library will have to continue to keep a hard copy of the information." ■■

ALA Spotlight Article

Starting Over: Current Issues in Online Catalog User Interface Design

Walt Crawford

Note: This paper was delivered as the keynote address at the LITA Preconference on Screen Design, June 27, 1991, in Atlanta. The paper was not prepared as a scholarly treatise and does not include citations for sources. Some relevant sources are cited in Walt Crawford, *Patron Access: Issues for Online Catalogs* (Boston: Hall, 1987); a few others may be cited in Walt Crawford, *The Online Catalog Book* (Boston: Hall, 1992).

What would your online catalog look like if you could start from scratch? Even if you can't, chances are that your next online catalog will offer extensive local control over the user interface. We're also gaining common models for interface design and a common model for command-driven searching, together with massively increased power on the desktop.

As we venture forth to new designs, we must be aware of existing knowledge about user interfaces, and we must concentrate on real users and their actual needs. We must also deal with conflicts and compromises, since our catalogs must exist in the real world.

What would your online catalog look like if you could start over? In other words, how would you design an online catalog if you could start from scratch?

Walt Crawford is a Senior Analyst at the Research Libraries Group, Inc., and is Vice-Chair, Chair-Elect of LITA.

These are reasonable questions to ask at this point, because you may be able to do just that, or something very much like it, in the near future. Where it counts—at the user interface—libraries of all sizes and varieties can expect to have more flexibility in the future, more freedom to put together a fresh design. That doesn't mean that public and academic libraries will be rolling their own complete systems, as very large universities have in the past. It does mean that future commercial systems should offer design possibilities far beyond what has been available in the past.

We're not quite there yet, at least not in most libraries—and most of us won't be there for a few years to come. Meanwhile, we are gaining more flexibility, and a few libraries are striking out with completely new designs. Computers continue to provide more power for less money; higher-quality displays continue to get cheaper; and the models are now in place to make a crucial leap forward in user interface flexibility.

As we venture forth to design completely new systems, we must be aware of existing knowledge about user interfaces. We must also be aware of the purpose of a user interface: snazzy design and clever features are pointless if they don't serve the aims of the library.

The first portion of this paper will concern current issues in interface design. First, I'll discuss the user base and the reasons that libraries have online catalogs. Then I'll go over some contemporary trends that are changing—and, I think, clarifying—the future of interface design.

After dealing with current trends, I'll go briefly over the well-established principles of user interface design, principles that are neither specific to libraries nor based on computer limitations. I'd also like to spend

a few minutes thinking about design principles that are specific to library catalogs. The third section of this talk will deal with conflicts and compromises, the messy realities that will always constrain designs.

Assumptions: MARC and Sufficient Power

I'm going to make two key assumptions. First, I'm going to assume that the next generation of online catalogs will be based on MARC-format records, some of them possibly enhanced and with other sources of information (citations, community information, etc.) added, to be sure. I don't see any economically feasible alternative; nor, to my mind, will that constrain design to any serious extent.

Second, future user interface designs will require substantially more computer power than past designs—but that power won't all be in a single computer. Until we get to the discussion of compromises, we should not worry about computing power or telecommunications bandwidth, at least when thinking about desirable future designs. It is no longer reasonable to begin with processing constraints before establishing a desirable design. First, decide what you'd like to do; then see if there is a realistic way to achieve it. After that, you can reduce your goals—but at least you'll establish a direction for the future.

The Mutter Test

Before going any further, let me mention a test that I find valuable in doing first-level hands-on evaluation of an online catalog; you might also find it useful. I call it the mutter test. It's simple: when you look for something in the online catalog, do you find yourself muttering under your breath? If a system passes—that is, if you don't start muttering—it may not be perfect, but it is at least adequate, for you, for now. There is only one passing grade, but there are gradations of failure, depending on how soon you start muttering and whether your muttering proceeds to grumbling, cursing, or barely restrained physical violence. If an online catalog engenders the fist-through-the-terminal response, it probably needs to be replaced, not just improved.

IF YOU COULD START FRESH . . .

Remember 1981, when the Council on Library Resources (CLR) sponsored studies of several online catalogs, some of which were really just fancy circulation systems? Remember the findings that most patrons just loved their online catalogs, even if they didn't use them? You would have thought the promised land of full user satisfaction was just around the corner.

That corner seems farther away now than it was then. That is partly because we know more and expect more—but it is also because today's users aren't the same as 1981's users. Before you can design your ideal user interface, you need to understand both words in that phrase: *user* and *interface*. Who are the users, and what is the interface?

Understanding the User Base

First, do you know who your users are? Do you know what they expect of the library, what they want from the catalog, and what they bring to the catalog? Every library is different, some dramatically so, but every library needs to consider its user base. I'll assert one universal factor, although I may be wrong: every user of a library catalog reads, at least once in a while.

Growing Base of Computer-Literate Users

Let's start with the biggest change in users over the past decade: computer literacy, for want of a better term. I'll bet that many of your users are comfortable with computers (mostly PCs), using them outside the library—and that is true for almost every library. That certainly wasn't true in 1981. The numbers vary from library to library, to be sure, but I'd be surprised if you found less than 25 percent of your users to be comfortable with computers, even in most public libraries. In academic libraries, of course, the percentages can be much higher, approaching 100 percent.

Every user who is comfortable with a PC is also comfortable with a keyboard. A growing percentage (but certainly not a majority) will even know how to use a mouse, if one is present. That comfort and knowledge gives you more flexibility in de-

signing a user interface. But users who are comfortable with PCs also expect performance from their systems. When you do something at the keyboard, you expect something on the display right now!

I'd even like to think that computer-literate users might know enough to distrust their first searches or even to distrust the online catalog, but that may be asking too much. Maybe next generation.

Dealing with Technophobes

Then there are the technophobic users. They may regard those dreadful machines as the ruination of the library: they knew how to find the right cards in those elegant wooden cabinets, and now they're lost. You can't really design an online catalog to deal with technophobes directly—although a well-designed user interface will at least keep the phobia from getting worse.

The one thing about technophobes is that they may not be whom you expect. Do you assume that your retired users are all hopeless when it comes to technology? Don't—given half a chance, many of the over-sixty set will take to computers quite as readily as anyone else. Actually, sad to say, there is reason to believe that the greatest pockets of technophobia are users in their forties and fifties. These are people who have always associated typing with secretarial labor and who think of computers as mysterious machines in sealed rooms, running their lives. In too many cases, they are also people who stopped learning shortly after they became successful and aren't about to start now.

The good news is that the percentage of users who are comfortable with computers is constantly increasing, even though there will always be a few technophobes. The bad news is that, some of the time, the technophobes are right. We all know that card catalogs do have some advantages over online catalogs.

"Tripping over a Librarian"

You do need to understand your users—all of them, not just the ones who cause trouble. And, of course, you need to understand the people who should be using your library and aren't, or the ones who use the library but never use the catalog. The latter group may know just what they're doing, incidentally—or they may be reacting to problems with your online catalog.

Ann Lipow has used a phrase that I think bears repeating, as a device for understanding your users, dealing with technophobes, and, in general, providing the missing piece for any online catalog. The phrase is "tripping over a librarian." In other words, users should find librarians on their way to the online catalog and should be aware that librarians are available to help. Those frontline librarians will learn more about real user needs and desires than any survey can reveal; they can also provide the expert service and advice that no catalog can.

Understanding the Application

Now, let's look at that second word, *interface*. A good online catalog is an interface between the patron and the library's collection. It may be more than that, but it must, first and foremost, do that job effectively.

Bulk of Use: Window to the Collection—Fast In, Fast Out

Most users in most libraries, most of the time, want to spend as little time as possible at the online catalog. They use it to get to the collection—to the books actually on the shelves in that particular location at that time. Most users, most of the time, don't want or need an exhaustive picture of everything that is available on a topic—they want something, and they want it now. Most users, most of the time, don't want to wait for material to come from some other location. Most users, most of the time, aren't interested in electronic access to full text.

We all know that is true (with the possible exception of special libraries). We also know that many users are impatient and unless they can understand, use, and get results from the online catalog within two or three minutes, they'll simply ignore it next time around, going directly to the shelves. On the other hand, if they can get in, get a call number and availability information, and get out fast, they'll be more inclined to use the online catalog when they need to know more—and they'll be happier with the library.

There is a lot of talk these days about the need to put users in touch with all the infor-

mation on a given topic, with the assertion that the systems are failing if users aren't made aware of all the possibilities. Not so. Librarians have never been able to inform users of all the possibilities; I dare say librarians don't really know all of them. And most users, most of the time, have neither the time, the need, nor the inclination to deal with all the possibilities. Librarians may suffer from information overload; you're neither obliged nor entitled to make sure that your users do.

In public libraries, of course, much of the time patrons aren't looking for information at all. If I go to my local branch library to see if something by Richard Condon that I haven't read yet is on the shelf, I don't care to know how many articles about Richard Condon are available. I know what I'm after. If I check the online catalog and suddenly realize that I need to browse more than one shelf area (because his novels are split between mystery and straight fiction), the online catalog's done more than its job—but if it takes more than two minutes at the terminal to find out what is available, I may never find that out!

Minority Use: Reference and Research

But, of course, we don't spend our energy designing for most users: that is obvious from many online catalog designs, and particularly from the tenor of almost all articles published in the area. Quick use isn't interesting use.

Instead, we focus on that important minority of users who are doing serious reference and research work and those users who have special needs not met readily by quick-check catalogs. That is fine; a lot must be done to meet the needs of advanced users.

What isn't fine, at least for most libraries, is to design systems that serve advanced users well and quick-check users poorly. If your online catalog offers wonderfully sophisticated prompting for Boolean searching and guides users gently through subject authorities and the process of searching your entire state's holdings as one, that is well and good—but if you need to go through seven or eight screens to get to the call number for a single item, you're not serving the majority of your users' needs.

In short: know your users—and recognize that most of them, most of the time, want to get into the catalog, use it briefly, and get on to the actual book or whatever.

CONTEMPORARY TRENDS

Now let's look at what is happening that should affect online catalog design over the next few years. I'd like to focus on four major trends that, among them, will improve familiarity for PC users, familiarity for experienced searchers, approachability for newcomers, and flexibility for system designers.

Common User Access: Familiarity for PC Users

First, there is what IBM calls common user access, or CUA: a common basis for user access, across hardware platforms and across applications. IBM didn't invent the idea, of course; proper Macintosh applications follow user interface guidelines that are reasonably similar to CUA, and the concepts go back to Xerox PARC in the early 1970s.

CUA designs are instantly recognizable to most PC users of any stripe. The top-line menu, one or more lines of environmental information below that, probably scroll bars on at least one border of the screen and, quite probably, one or two additional information lines at the bottom of the screen.

CUA doesn't help at all when a user has not used a program along these lines—but it provides instant familiarity if a user has. Today, that includes every Macintosh user, even though they need that "press Alt" note if there is no mouse. It also includes every Windows user, every PC Tools Deluxe user, most Quattro Pro users, every Ventura Publisher user, and every user of Microsoft Works or Microsoft Word 5.5. Add those together, and you have millions of people familiar with these concepts, right now. There will be more in the future. They are still a minority of library users, but a growing minority.

Keyboard Norms

If the online catalog uses PCs for its user interfaces and has that top-line menu, many users will assume that F1 brings up

Help—and, less consistently, that F10 activates the menu (as an alternative to Alt or a mouse). Those two assignments establish rudimentary keyboard norms. Less commonly, F3 should leave an application—but for dedicated online catalog terminals, that may not mean anything.

Keyboard norms have always been a sore point in online catalog usage and in PC software. CUA doesn't provide a total solution, but it does provide a start. I'd like to think that future online catalog interfaces will also make better use of common self-defining keys on PC keyboards; for example, PgUp and PgDn should do what they say.

Pointers and Menus

It is well to remember that the top row of menus doesn't imply that a program is always menu driven. It does, however, imply that menus are available when you need them. You can expect the menus to drop down with single-letter choices, some of which lead to secondary menus or, quite possibly, dialogue boxes.

That is all well and good, but it is also frequently the slow way to do something, particularly if what you want to do is prepare a search or move through a multirecord display. If you're preparing a search, you're always going to wind up with a dialogue box in any case (you can't search without providing text). If you are moving through a large set of results, the last thing you need is numerous menus.

The best PC programs offer many shortcuts for the most commonly used options so that you don't have to deal with menus most of the time. So should user interfaces, although here the alternative may be a bit different. After all, the online catalog user probably isn't editing within the display screen; is there any reason that commands shouldn't be recognized and acted on?

Common Command Language: Familiarity for Searchers

We'll see more catalogs that offer a familiar interface for people who have used contemporary PC programs. I think we'll also see more catalogs that support a single command structure, and you probably already know what that command structure is. Common Command Language (CCL), NISO Z39.58, will provide familiarity for experienced searchers and for those who use more than one online catalog—and it will provide the language of choice for frequent online catalog users.

Universal Alternate Mode

I am not saying that future catalogs will require users to learn CCL; after all, common user access implies that you can get started without knowing a command syntax. Some catalogs will provide excellent access without ever requiring users to learn a syntax; some do right now.

What CCL should provide is an increasingly universal alternate mode, a way of moving among catalogs and getting more out of catalogs. Indeed, we might see future catalogs that depend on CCL for advanced modes of access. Remember: most users don't require sophisticated access. Remember also that CCL is easy to learn and easier to remember.

Using CCL in a Friendly Environment

I envision at least one style of user interface that effectively teaches the user CCL as he or she traverses menus. Say there is a command line (maybe not called that). Say that the command line fills itself in as you make menu selections. You never *need* to use the command line; between menus and dialogue boxes, you can get everything done elsewhere. Then again, that filled-in command may be echoed back to you—after all, there has to be something that shows how you got where you are. Remember signposts; we still need them.

The casual user may simply ignore that text, even though it will all make pretty good sense; CCL is, after all, essentially a simplified form of declarative English. But the experienced user, or the frequent user who has grown tired of menus, may notice that there is a command line with English-like text in it, and might even try typing something.

There are other routes: for example, a catalog with its own style of command entry could have a function key to call up a CCL dialogue box. In any case, adding CCL provides ready access to a growing

number of users—like CUA, but for a different user base.

I am a California native. To me, CCL is already second nature; after all, SPIRES/BALLOTS/RLIN pretty much started the whole thing, and other West Coast systems have followed that lead. They may not be pure CCL, but if you know one of them you'll have a pretty good shot at the others—and that is the whole idea.

Context-Sensitive Help: Approachability for Newcomers

We have one trend, CUA, that will make some future online catalogs more accessible to PC users of all stripes. We have another trend, CCL, slightly conflicting with the first, that will make them more powerful and familiar for multisystem users and researchers. But what about the others, particularly those coming to this catalog for the first time?

Good opening screens and good context-sensitive Help should play a role, although nothing will replace the nearby library worker. Help has always been with us, but there is a fairly new way to organize Help that can make it far more powerful.

Hyperlinks

It's the "h word": hypertext or, more specifically, hyperlinks within a Help system. The context-sensitive Help that you get initially may have highlighted words and phrases that will lead to additional Help on certain topics, with some reasonably clear way of getting back where you were. Of course, any good Help system will also allow you to take other actions directly from the Help screen or window (and, indeed, a Help window may make more sense than full-screen Help).

Hyperlinks work best when a mouse or other pointer is available, and they can be more subtle and less annoying if the screen supports color, graphics, or both. It is unfortunate but true that hyperlinks will be most helpful to people who understand PCs, since they're more likely to see how the links can be explored. (Naive users may find hyperlinks totally befuddling.)

Remember that context-sensitive Help is not the same as online tutorials, and there is a place for both in a good online catalog. Help itself should say something related to the current situation, whatever that might be. Tutorials should offer information about the online catalog, possibly at various levels of specificity. In CCL, the distinction is between Help and Explain, where Explain brings up tutorials. With hyperlinks, minitutorials may come up through highlighted words in context-sensitive Help; those minitutorials may further lead to extended tutorials—or, using CCL or function keys, users might go directly to tutorials. But enough about Help for now.

Moving the Power to the Desktop

Last of the current trends, and underlying at least two of these other trends, future online catalogs will take advantage of the awesome increase in distributed processing power. In other words, more of the processing will move to the individual desktop and away from the central system. That process may be slow; most libraries can't amortize systems at reasonable rates, and some of you are doubtless still buying dumb terminals or using workstations to emulate dumb terminals. But the process has begun, and it will continue; I'd like to think that dumb terminals will decline very rapidly in the next year or two.

Earlier I mentioned NISO Z39.58, still a draft standard. The migration of power to the desktop will be based on another NISO standard, this one fully adopted and being revised: Z39.50, the Open Systems Interconnect (OSI) standard for information-retrieval service.

What Z39.50 and related standards do now and will do in the future is to provide a well-specified channel for sending requests such as searches in one direction and sending responses such as result sizes and actual records in the other direction.

What it means for future online catalogs is both straightforward and very exciting. I believe we'll see systems where the central computer (whatever its size) does almost nothing but process searches, sort results, and transmit results. It won't know from user interface, and it won't care; it will accept well-formed queries and requests, and it will respond with set numbers, result sizes, and MARC records.

Everything else will happen on the desktop—or, for dial-up use, perhaps on another computer that serves as a multiuser virtual desktop. Everything else includes menu handling, command parsing (or command creation), all display logic, including conversion from MARC to labeled or citation-style displays, and so on. Context-sensitive Help will rely on clues from the desktop; the actual Help and tutorials may be strictly on the desktop or may involve some back and forth. The desktop system, the user terminal, won't process the search and won't address the database directly—but it will do just about everything else.

That means power, flexibility, and openness. It means that the same system can have pure-Macintosh user interfaces on Macs, pure-PC user interfaces on PCs, Spanish-language user interfaces where they're useful, and so on. It means that staff and vendors will be able to experiment with new interfaces without any change in the central computer. It means that libraries may be providing software for home computer users and other remote users—and it means that the same user-interface software may be useful for many different remote systems.

It may even mean different kinds of competition in the online catalog market. Some systems have more efficient searching algorithms and database managers than others, and some systems have better user interfaces than others. What if you could choose one vendor for your central system and another vendor—or more than one other vendor—for your user interface? Why not?

Well, those are a few trends that I think will be important in this decade. There are lots of others, but these may be defining trends, as they provide the framework for other improvements. Taken together, they can make enormous differences.

REMEMBERING KNOWN PRINCIPLES

What is interesting here is that you really may be able to start fresh. What is dangerous is that some libraries and new vendors will do so without paying attention to known principles of screen design. The bibliography you receive will lead you to much more information on those principles, but it is worth noting a few of them here.

Most of these principles have little or nothing to do with libraries and little or nothing to do with character-based user interfaces. They are principles based on studies of short-term memory, readability, and so on. They're not gospel, exactly—but I've never seen systems that ignore them work as well as systems that follow them.

Density and Sparseness: The "30 Percent" Rule

First, there is the "30 percent" rule, which is related to displays in general. This rule says that no more than 30 percent of the screen should be filled with characters: in other words, the density of information should be 30 percent or less. The ideal, supposedly, is 15 percent. Effective displays are open displays, with lots of empty space used to clarify the important information.

A related principle, based on many studies of readability, says that lines of text should be no more than sixty to sixty-five characters long, whether the text is proportional or fixed. There is also the general sense that multiple-record displays that use more than one line per record should have blank lines between records.

Put all of those together, and you'll find that 30 percent is a pretty good target; indeed, screens with more text seem crowded and hard to deal with. For most record displays, the 30-percent limit will fall out from other design aspects; for Help and tutorial screens, it is a good limit to keep in mind. It is also good to think in terms of self-contained single-screen Help and tutorial texts, even though scrolling may be easy to do. After all, six hundred characters is a hundred words; that's a good length for a terse single-topic message.

"Seven Plus or Minus Two": Limiting the Choices

Another principle, well established through extensive testing, is that people can only deal with five to nine options at a time, or five to nine discrete pieces of information. It is the "seven plus or minus two" rule, and you should consider it a sound rule for interface design. Basically, a typi-

cal person can take in a set of seven to nine options at a glance, where more options will require more thought and slower reaction.

That means that single-line displays should probably only include nine records, even though there is room for fifteen or sixteen. Some current catalogs do precisely that; the extra blank space may seem odd, but it is much easier to focus on the small set of records.

It also means that a good user interface will have no more than about nine things going on at once and no more than nine simultaneous options. Limiting your design to nine options at a time forces coherent organization of options—including that all-important "other"—but frees the user to choose immediately, at a glance. Faced with a dozen or more choices spread around the screen, no matter how cleverly arranged, the user must read through the choices one by one—an inherently slower process. I tend to mutter when faced with more than nine choices at once.

From what I have seen of some contemporary interfaces, this may be the most frequently forgotten rule—and based on all of my own experience, I would say it continues to be a critical one. Disneyland succeeds so well partly because there are never more than three choices at a time. You don't have to go that far, but learn to organize in groups of no more than nine.

It is never that simple, of course. A pull-down menu may well have more than nine options, but good ones will always have options grouped—and there will never be more than nine within a group or nine groups. A selection list may have hundreds of options—but you'll find that most good ones only show eight or nine at a time. And, of course, you may have six options on a menu bar, a dozen options (in smaller groups) on a pulled-down menu, and several more on a side submenu or dialogue box. But that is OK: you're never actually requiring the user to deal with more than nine simultaneous choices.

Try it yourself. Look at a long list of options. Then organize that list so that there are never nine choices at once. I think you will find that you can deal with the latter much more rapidly.

I have spent a lot of time on this, but it is perhaps the most well-established rule in the field—and one of the easiest to overlook in the excitement of building a new interface. Don't disregard it.

Navigability and Control

The next principles work together: the user should always be in control, and the user should be able to navigate freely within a system. Both are important, particularly in an online catalog. Too often, we feel that we are serving computers, rather than that they are serving us. This is totally unacceptable in an online catalog.

Don't surprise the user, and don't do something the user has not asked you to do without letting the user know about it. That is user control, and it is vital. It is not really limiting, but it is easy enough to violate. For example, if a user searches for an author or subject using a form that isn't used, it is perfectly appropriate to bring up a list of headings, showing "This library uses *blatzo*" and putting a prompt on the blatzo line for an immediate search. It's also appropriate, if you can't display such a list, to display a message saying "This library doesn't use the term *illiterate*, we use textually *disadvantaged* instead. Would you like to see those records?"

It is inappropriate, however, just to change the search, particularly when the catalog then displays feedback, incorrectly saying that the user searched by the correct term. It is equally inappropriate to throw away portions of the user's search without saying so, or—to use a real example that should now be obsolete—to turn the user's textual search into a derived key with predictably bizarre results without saying so.

Free navigation is an important principle for future catalogs, one found in most (but not all) contemporary command-driven systems and few (but some) menu-driven systems. The principle is simple: you should be able to take any legitimate action at any point. The clearest case is Help screens: once you have learned how to do something, you should be able to do it *while you are still looking at the instructions*. Requiring the user to escape out of Help before issuing the new command is just plain rude, albeit easier to implement.

The same goes for requiring users to return to a multiple-record display in order to go from a long display of record 3 to a long display of record 4—or, worse yet, requiring users to back up through several levels before doing something entirely different. I'll give a real example here. If I am a knowledgeable user who can't be bothered with *LCSH* in all its glory, I will come in with a good title word search to find one item on the subject I want. I may then go to a long display for that item to find a subject, then use that subject to find more items. That is common behavior, and it is intelligent behavior. But a surprising number of systems simply will not let me issue a subject search directly from that long screen while I'm looking at the record. Why not? Well, because it isn't one of the current options. . . .

Of course, the best solution in this case is to let me highlight the subject and call up a related-item search. Let the computer figure out what index to use. Some systems have been doing that, although usually in a slightly different sequence, for more than half a decade; it is a fine idea, and one that should be replicated in every future catalog. (Now, I'll admit that is one where I think pointing a mouse cursor at the subject entry and double clicking, while inadequate as the sole methodology, is perfectly sensible as a shortcut.)

Clarity and Consistency

Any good user interface must be clear and consistent. Clear interfaces offer distinct options (whether commands or menus), don't clutter up the screen with too many choices, and use carefully written text and labels. Clear interfaces avoid library jargon (and if you don't think *imprint, collation, uniform title,* and *tracing* are all pieces of library jargon, think again). Clear interfaces use at most one or two highlighting techniques and use them for straightforward purposes.

Consistent interfaces always use the same words to mean the same things. Linguistic variety is a great thing in a novel or a long article, but not when you're offering fast information on a mildly arcane topic. I don't know whether the best word for one of those 6xx fields is *subject, topic,* or *about*—but I do know that any one catalog should always, always use the same term. And yes, that does mean checking the text of every screen, every Help message, every tutorial—not only for appropriate tone and clarity, but also for pure, simple consistency.

Consistent commands also improve a user interface—and the consistency should be apparent. Don't insist on distinctions that only make sense to librarians or system designers. If two functions that are really slightly different don't overlap and appear to be essentially the same from a user's perspective, use the same name for both of them.

There is a lot more to clarity and consistency than these comments, but this is a start.

Forgiveness

Consider forgiveness as part of your user interface. If an online catalog design can deal with minor problems in user commands, it should—either taking unambiguous actions or prompting for ambiguous cases. I'm of two minds about spell-checking; in a large academic library, I suspect it is slightly worse than useless, but it could improve forgiveness in smaller and less multilingual libraries.

Other aspects of forgiveness get tricky and involve some difficult decisions. You should keep forgiveness in mind, but remember that the user must stay in control. The system shouldn't be forgiving to the point of essentially ignoring what the user actually did.

But there is one key aspect to forgiveness: while the user may do something that the system either does not understand or cannot process, the user cannot make an error as such. That is, the word *error* should never appear in online text!

"Obviousness"

To wrap up this partial set of known principles, consider the need to be obvious. What does that mean? Basically, that the way a catalog works should be pretty obvious within a few minutes or at least obvious enough that a casual user can complete a casual search. That is not a terribly difficult criterion, but it is an important one. Let's

face it—you're not going to get every user into a training session, and infrequent users would forget the training in any event. An obvious interface should just, well, work.

THINKING ABOUT PRINCIPLES

As we move into areas of conflict and compromise, there are two aspects of online catalog user interface design that have no easy answers but that definitely deserve some thought. They are the metaphor established in a good user interface and the question of coherence between the interface and the underlying engine.

Familiarity: The Right Metaphor?

You probably know the usefulness of a good metaphor in a new system. If the users can relate the new system to something they are already familiar with, they will be more comfortable.

Danger of the Card Metaphor

But that familiarity can also be a trap. I used to shudder when someone wrote or spoke about "online card catalogs"—and find it horrifying that some people have suggested using HyperCard or another graphic methodology to make an online catalog look sort of like a card catalog.

What a terrible idea! A good online catalog has very little in common with a good card catalog, except that both show bibliographic information and lead the users to actual books and other items. But a good online catalog offers more kinds of access than a card catalog, should offer more resources in the future than a card catalog, and should offer status information, which a card catalog can't do. If users buy into the card metaphor, they'll ignore some of the most valuable searching methods and other uses of the online catalog.

That isn't the worst problem, though. The worst problem with the card metaphor is that an online catalog is a terrible card catalog, and there is no way to fix that. The process of flipping through cards offers a form of fast filtering and information retrieval that really isn't possible on a screen. And the more you make the online catalog look like a card catalog, the more patrons will be aware that they've lost something.

Labels: Help or Hindrance—and What Labels?

One of the things an online catalog can provide, which very few card catalogs provided, is labels for information. But should it—and what should the labels be? What a question, right? Of course you should label fields in a bibliographic display; that way, users will finally understand what all that information is.

Well, yes—or maybe not. The anti-label side is really quite simple: you can usually show a fairly complete bibliographic record, including subjects and notes, on a single screen in a traditional bibliographic format—and you usually need two screens or more when you add labels and appropriate spacing.

And, of course, if you use labels, you need to determine what those labels are. Not all that easy. I have yet to see a name for the 265 field or the 300 field that causes me to jump up and say, "That's it!" And I'm not sure there is a single best name for 6xx fields or 7xx fields or some others. Arcane fields, particularly coded fields, are very difficult to label—but they may be inappropriate for display in any case.

Much of the research that has been done asks the wrong questions, I think. The question is not really whether a user can identify (that is, name) a given piece of information. The questions are whether the user got the information he or she wanted, how quickly he or she got the information, and whether the process was satisfactory. You can be certain that some users want to know some of the information in 300 fields, particularly for nonbooks, just as someone looking for a version of the *Missa Solemnis* will certainly want to see the 511 (Performer) note. But you can be equally certain that most users really don't care what the information is called, just as long as it is visible.

I will be honest: this one drives me nuts. I can give you lots of advice about how to design a good labeled display—for example, that the labels should be right aligned and either all capital letters, reverse video, or otherwise highlighted, and that the fields should be left aligned. But should the default single-record display be labeled? I just don't know.

Carrying Through on a Metaphor: Extendability

One more thing about metaphors and overall design: if it is not extendable, it is a mistake. If your online catalog succeeds, you will add to it: possibly citations, possibly community or campus information, possibly gateways to other catalogs, possibly image-retrieval systems. You're also likely to add new functionality as time goes on.

It is one thing for the initial user interface to establish a clear, consistent metaphor and model. It is quite another to be able to add to the system without disrupting that clarity. Some designs are open-ended; some are not. Go for the open-ended designs if you have any choice in the matter.

Coherence: Reflecting the Underlying Engine

Here is a design criterion that should not need saying. The user interface should not fight with the user. This seems obvious, and yet haven't you used systems that required a return key in one place and accepted a direct keystroke in an apparently similar place? What about systems that appear to accept commands, but won't actually execute workable commands unless you're at the right place in the system— for example, that won't accept a new search while you're looking at a record, but require you to go back to a search screen first? Now, what about the ones that not only won't do the logical thing—but actually recognize what you want to do and *tell* you how to go about doing it?

Unfortunately, all of those are real examples that I've used in the past few months and are examples of currently available systems. They shouldn't be that way; it is particularly appalling to have a system that won't accept a consistent command in one situation *even though it recognizes the command and knows what it should be.* As you can guess, such systems fail the mutter test very rapidly. "Give me a break!" is the appropriate response.

Frustrating problems like this can happen for several reasons, but one reason can be a lack of coherence between the user interface and the underlying system. In at least one instance, I have seen an interface that appears to accept commands—but there is a strong indication that it is built on top of a menu-driven engine. And so, of course, some commands (not being on the current menu) just cannot be executed; the user interface can only tell you how to get to where you can proceed.

CONFLICTS AND COMPROMISES

Five years from now, such systems as just cited could be long gone, at least it would be nice to think so—although, fiscal realities being what they are, it seems improbable.

Time to discuss conflicts and compromises. That brings in some other current trends not covered earlier. If you thought the material above was something of a smorgasbord, well, this part goes even farther afield. There are seven topics to cover here, some of them already mentioned briefly, and I can't think of any way to fit them all together, except that they're all aspects of contemporary user interface design.

Power and Confusion: How Many Indexes and What Indexes?

First, how many indexes should a good online catalog include? I have heard and seen answers all the way from one (too few) to several dozen (far too many). Equally important, what indexes should they be, and how should they be arranged? You could argue that index choices aren't part of the interface—but of course they are, for a couple of reasons. One important question is how multiple indexes are presented and supported; another is whether the number of indexes visible to the user is identical to the number of actual indexes.

I am willing to argue for a total of five basic indexes to support very powerful catalogs, possibly with a larger number of more refined indexes available to those who need them and know about them. What five, and how would they work? Three internal phrase indexes: one for names (all names, personal and corporate, including 600–611 and 800–811 fields); one for topics (all 6xx and some 7xx fields); and one for titles (24x, 4xx, and 8xx, and title subfields elsewhere). An internal-phrase index takes phrases as arguments, but the phrases can appear anywhere within a field. Thus, for example, "fin title online reference serv"

will retrieve the book *Managing Online Reference Services*. Internal-phrase indexes will function as word indexes, although badly. And, by using double quotes, those who want leading-phrase retrieval can have it. As a fourth index, a word index combining all indexable text fields: big, sloppy, but effective. Finally, a codes-and-numbers index combining all reasonable coded retrieval fields such as ISBN, publisher's number for sound recordings, ISSN, CODEN, and so on.

Is that set optimal? I don't know. Does any current catalog use it? Not that I know of. Would it work well? I believe so. What is your optimal list—and why?

Pointers and GUIs: Obvious for the First-Timer?

But what about graphic user interfaces (GUIs) and mice in general? Do they make sense for general use? Yes and no. The fact is that the pure Macintosh interface fails the two-minute test: someone who has never used it cannot sit down to a Macintosh without any instruction and be effective immediately.

Which only means that there must be some sort of signage that tells you how to get to an opening screen (and that method *must not* involve the mouse, which isn't an inherently obvious device) and that the opening screen must give you enough information to keep going—again, without requiring the mouse.

If theft and space are not problems, adding mouse sensitivity to the interface makes excellent sense; many users will understand it, and even more will find that it speeds up some operations. Theft may well be a problem—but you can get keyboards that have built-in track balls, reducing that particular problem. Don't *require* the mouse, not unless you are on a campus where everybody uses Macs. Just like alphabetical order, it is only obvious once you understand it.

Color and Color Blindness: The 8-Percent Factor

So far, I haven't mentioned color. Five or six years ago, it was easy to advise against color; not only were there few, if any, proven effective uses for color in a user interface, the color monitors available were either very expensive or much less readable than monochrome monitors—and, for that matter, the colors available were frequently few and fairly garish. The second factor has gone away; today's $300 displays can provide excellent color from a vast palette, with textual clarity equal to most monochrome monitors.

That removes one of the major barriers to using color. It does not automatically make color worthwhile, and I am not sure that a case has been made one way or the other. I will say that color needs to be used thoughtfully and subtly in user interfaces, particularly when it comes to red, perhaps the only color with a generally understood significance. One thing can be said: while color may be useful to enhance a display, it must not be used as the only means of conveying critical information. Why? There must be at least one or two men in the audience who know the answer, even though male designers are as prone as any to make the mistake. At least 8 percent of men suffer color blindness and simply won't get some of the color-coded information. At worst, some combinations of color and background may be invisible to some users: for example, and surprisingly, the specially colored fifty-five-miles-per-hour indicator on some car speedometers is invisible to some men.

Am I saying to forget about color? Not necessarily. There is no getting around the fact that a well-designed multicolor text application, running on a high-resolution color monitor with good glare and reflection control, is just more pleasant to use than the same application running in black and white. More productive? I'm not sure, but after more than a year of using VGA, I'm not ready to dismiss the possibility.

On the other hand, that is mostly an issue when you'll be using a system for hours—and very few online catalog users do, or should, use it for more than ten minutes at a time. Color can be used to attract, but I'm not at all sure it is needed. I am sure that, if and when color monitors become the natural choices, color must be used thoughtfully.

Effective Help—and Unlikely Use

When it comes to conflict, one classic

area is online Help, including both context-sensitive Help and online tutorials. On the one hand, every good user interface requires thoughtful, effective, compact Help screens, with clear ways to get at more information and with clear ways to execute the next command from the Help screen. It takes real effort to write clear, consistent, terse Help screens—and it is important.

But you cannot rely on online Help and tutorials to inform users or get them out of trouble. If there is one thing we do know about almost all user interfaces and their Help systems, it is that most people will not ask for help even when they need it and the function is clearly labeled.

You probably know that from your own PC applications, if you're honest about it—I certainly do! And log analysis within online catalogs has affirmed what we probably already knew: Help screens are among the least-visited screens in the catalog.

I don't know why; it is partly human nature, I suspect, just like those who will spend an hour driving around, hopelessly lost, rather than pull into a gas station and ask for directions. Maybe people don't want to be seen looking at Help screens—or maybe they just get into trying to correct their own problems without thinking that the system can help.

Will it get better? Possibly, although probably not much. In any case, you do need the printed one-page guides to the catalog, you do need flip charts (if you have room), and you do need, most of all, readily available library staff. But you also need effective online Help, even if the people who need it most may not use it that much.

Arcane Elements: Completeness and Confusion

Jumping to an entirely different topic, what do you do about displaying the more arcane elements of a bibliographic record? How do you strike a balance between completeness—the desire to show everything that is there—and confusion—the problems that arise because elements aren't self-explanatory?

I think notes belong on a long display. Subjects and added entries belong on anything but the briefest display. Repetitive though it frequently is, the statement of responsibility probably belongs on bibliographic displays.

But not the true arcana, unless (1) it can be clearly explained and (2) there is some specific reason to want to include it. Country of publication? Why bother? It is generally present or can be inferred from the publication information; why display the codes? What about language? Well, if the user can read the language that the title is in, the user can probably recognize the language. If the user cannot read it, just how much does the user care what language it is in? Personally, if I am looking for information, I am not too concerned with whether the book represented by a catalog listing is in Vietnamese, Albanian, or French: the chance of my getting anything out of it is terrible in any case.

Most coded elements, excepting standard numbers, should not be displayed except in MARC-tagged displays. As for textual fields—well, some patrons definitely want to know how many pages are in a book when they're choosing which one to read. There is much valuable information in notes; indeed, for many items (such as classical sound recordings and anthologies), the notes may provide key information for selecting items.

Other Catalogs

Then there is the question of multisite access. Your local library's user interface will soon be able to serve as a gateway to other catalogs, ranging from your countywide union catalog or multicampus catalog, through regional or statewide catalogs, to resources in other states and other countries. With Z39.50, it is quite possible that such access can be transparent to the user.

Fine, but not if the poor user is forced to deal with all those resources as though they were part of the local catalog! Not to harp on this, but the library has no business forcing data overload on its users—and forcing users to consider external resources is a way of forcing data overload. Once again: most users, most of the time, want something now, which means something that is on the shelves in the building they're currently using. That is not only human nature, it is a rational desire.

A library catalog should always start out

as a catalog to the collection of that library—with sufficient pointers, if you will, to let people know that they can go much further if they wish. But don't force it all on people; that is a serious disservice.

Serving the Remote User

Finally, what about the remote user? How will that user be affected by these glitzy new user interfaces, particularly when the user interface itself migrates to the desktops within the library?

Good questions and tough ones. Several answers come to mind, none of them complete. One answer: dial-in lines can connect to a microcomputer that is functioning as a multiple-user-interface front end, albeit probably with a less powerful version of the user interface.

Another answer: dial-in and Internet/Tymnet connections can be based on user interface software that the library provides. That is how Prodigy works; while Prodigy may be a mess, it is a clever mess all the same.

Yet another answer for the medium term, albeit one that will give user-services librarians absolute fits. As Z39.50 becomes more widely established, people may be able to obtain their own user interface programs that can then be used with any accessible Z39.50 server. Which is one good reason that your own interface software should include the name of the library in the menu bar; that way, when you get a call for help that doesn't make sense, you may be able to determine that the user isn't using your own software.

CONCLUSIONS

No Perfect Answer

First, as you are pondering your next online catalog interface, remember that there is no single perfect answer. No online catalog design is the best design for all users in all libraries at all times. For that matter, I am not sure that the concept of an ideal design *for a specific library* makes much sense. Ideal for what? For whom?

A successful user interface satisfies the fast lookup needs of the majority of users, while providing enough power to satisfy the deeper needs of the minority. It gives ready access to call number and status information without plowing through many screens, but also lets people see all the information they can use when they need it—and lets them decide what they can use, rather than restricting information to what some study shows will satisfy 80 percent or 90 percent of users. Aim for success; aim for improvement. Aim for coherence, clarity, and excellence. Don't worry about designing the perfect interface: you won't, and that is not a problem.

Teaching versus Preaching

Second, remember that a user interface should teach, not preach. What that means is that a user interface should make it easy for a user to learn more about what is available—but should not force the user through extra steps to be sure that the user understands proper bibliographic protocol.

That is another way to look at the issue of access to outside resources. It is useful to teach the user about them, with an appropriate line in the one-page reference and maybe, just maybe, a suggestive prompt when a search yields no results (or yields results, none of which are currently available). That is teaching. When you force the user to make an explicit selection of the local collection, or even worse present the user with a broader result to be plowed through, you're preaching. Don't make a religion out of the online catalog; it is just a tool.

Defining Failure

Which leads right into the third point: be sure you know what constitutes failure. If a patron looks for a work and fails to retrieve it—even though it is there—because the system is badly designed, does not accept synonyms, or otherwise gets in the way, then the online catalog has failed. True enough, and very difficult to deal with.

But if the user keys in a search that only retrieves a fraction of what might be available on what appears to be desired topic, and the user goes off to get a book from that partial list, *there is no reason to assume failure.* Yes, maybe another book would have been better; yes, maybe there are articles that would provide more current information. And maybe an intervention by a librarian or some sort of expert user information system would lead to that better

information. But just as likely, it really doesn't matter. The user wanted some general information, has gotten it rapidly, and is both happy and successful.

The catalog fails—the library fails—when the collection is not represented or when material that is present is not retrieved through reasonable means. The catalog does not fail when users don't find out everything there is to know about a topic.

No Substitute for Service-Oriented Librarians

Finally, the most important parts of any successful user interface are not computerized at all. They are the true expert systems to help users gain access to the collection, and they go by the title "librarian." Service-oriented librarians combined with a mediocre online catalog will provide better real user service than a first-rate catalog with no librarians in sight. Fortunately, I do not see that changing any time soon. Unfortunately but realistically, it means that local users will always be served better than remote users. The library continues to be an important physical location; that is not such a terrible thing. ■■

Statement of Ownership and Management

Information Technology and Libraries is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611-2795. Annual subscription price, $20. American Library Association, owner; Thomas W. Leonhardt, editor. Second-class postage paid at Chicago (Cook), Illinois. Printed in U.S.A. As a nonprofit organization authorized to mail at special rates (Section 423.12, *Domestic Mail Manual*), 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—the June 1991 issue.) *Total number of copies printed:* Average, 7,401; Actual, 7,424. *Paid and/or requested circulation:* not applicable (i.e., no sales through dealers and carriers, street vendors and counter sales). *Mail subscriptions:* Average, 6,719; Actual, 6,783. *Total paid and/or requested circulation:* Average, 6,719; Actual, 6,783. *Free distribution by mail, carrier or other means, samples, complimentary, and other free copies:* Average, 166; Actual, 176. *Total distribution:* Average, 6,885; Actual, 6,959. *Copies not distributed: office use, left over, unaccounted, spoiled after printing:* Average, 516; Actual, 465. *Returns from news agents:* not applicable. *Total (sum previous three entries):* Average, 7,401; Actual, 7,424.

Statement of Ownership, Management and Circulation (PS form 3526, Feb. 1989) for 1991 filed with the United States Post Office Postmaster in Chicago, September 20, 1990.

News and Announcements

FLICC/FEDLINK Operates under First Comprehensive Bylaws

With the official signing by Librarian of Congress James H. Billington on July 16, 1991, the Federal Library and Information Center Committee (FLICC) adopted its first comprehensive bylaws.

The bylaws document stipulates that FLICC will continue its roles, coordinating cooperative activities undertaken by the more than 2,400 federal libraries and information centers across the United States and at U.S. facilities around the world, providing a forum to make recommendations affecting this community through the librarian and promoting the application of new technologies by the community and improved access to information through such activities as the services provided by FEDLINK (Federal Library and Information Network).

The original chartering document for the Federal Library Committee—FLICC's name before it was reconstituted in 1984 to include information centers—was issued March 23, 1965. The original committee was established by L. Quincy Mumford, then librarian of Congress, at the behest of federal librarians with the cooperation of the Office of Management and Budget (OMB), then the Bureau of the Budget.

Even as the bylaws formalize practices and procedures that have evolved since FLICC's establishment, the bylaws establish an updated organizational structure. Here is how FLICC will operate under the bylaws:

- The librarian or his designate chairs FLICC, which is composed of fifty-seven federal agency members—thirty-two appointed by permanent federal agency members, fifteen elected by FEDLINK voting members, nine elected as rotating members by the permanent members, and one serving ex officio by virtue of election as chair of the newly established FEDLINK Advisory Council (FAC).
- The FLICC chair designate also chairs the new FLICC Executive Board (FEB), with the aid of a vice-chair chosen annually from the FEB's ten members—three appointed by the FEB chair, six elected by FLICC members, and the FAC chair.
- The FAC, one of six standing working groups in FLICC, will consist of nine elected FEDLINK members serving staggered terms. The three receiving the most votes among the five chosen annually by FEDLINK members as its representatives to FLICC will join FAC as the three who have completed their three-year terms depart.

The FLICC Executive Board will have a central role in such areas as establishing and monitoring FLICC's working groups, providing oversight on FLICC programs and budgets and on FEDLINK operations and initiatives, and recommending policies, objectives, plans, and annual budgets. As the lead working group, FEDLINK and its FEDLINK Advisory Council will aid the FLICC executive director in FEDLINK operations; provide direction to the FEDLINK OCLC Users Council; submit to the FEB for approval initiatives for new FEDLINK services, program objectives, plans, and the annual FEDLINK operating budget; and request initiation of appropriated agreements for FEDLINK with various services, systems, groups, and agencies within the authority granted by FEB. The five other standing FLICC working groups are: Finance, Policy, Nominating, Membership and Governance, and Education.

Designed to regulate FLICC's affairs and govern the activities of its members, the bylaws are expected to be fully implemented by January 1992, completing a year of transition planned by FLICC/FEDLINK's members and management and by Levering, who was appointed executive director in November 1990 after serv-

ing twenty months as acting executive director. ■■

Innovative Interfaces Offers Art Reference Database

Color images of Renoir's *Odalisque*, Monet's *Vase of Chrysanthemums*, and other renowned works of art can now be retrieved and displayed with the INNOPAC online catalog. Innovative's Art Reference Database includes more than 1,700 images of paintings, sculptures, drawings, and prints from the National Gallery of Art.

Library users search the INNOPAC Art Reference Database by artist, title, keyword, style, and medium. Users can also combine search terms with Boolean logic. For example, one might search for all works in the Impressionist style and in the oil medium. When a record is retrieved, INNOPAC displays a text description on a public catalog terminal and simultaneously displays the associated visual image on a color videodisk monitor. The Berkeley Public Library was the first library to install the Art Reference Database.

The Art Reference Database software is also available for the development of local databases. Libraries can create digital images by scanning photographs, maps, slides, drawings, or other visual materials. The images can then be linked by frame number to MARC records with text descriptions of the images. ■■

NOTIS Ships Z39.50 Software to Development Partners

NOTIS Systems has shipped PACLink, its Z39.50-based linkage software, to Indiana University and the State University of New York graduate centers. The product will be tested in Indiana by Indiana University (Bloomington) and Indiana State University (Terre Haute). SUNY-Binghamton and SUNY-Stony Brook will test the software in New York. The first general release of PACLink is scheduled for the first half of 1992. PACLink permits a library user at a local online public access catalog (OPAC) to make a seamless connection to an OPAC at a remote library. The user simply selects from a menu and searches the remote OPAC in exactly the same way as the local OPAC. Bibliographic data retrieved from the remote OPAC displays in exactly the same way as the local OPAC.

PACLink is based on the NISO Z39.50 standard, Information Retrieval Service Definition and Protocol Specification for Library Applications, and utilizes client/server architecture. Initially, connections between PACLink sites are via the TCP/IP-based Internet. ■■

CLR Names Julia C. Blixrud to Program Officer Position

Julia C. Blixrud, formerly section head of the National Serials Data Program, Library of Congress, has joined the Council on Library Resources as program officer, where her duties will include grants management and coordination of the council's proposal review process. She will assist the council's president, W. David Penniman, in carrying out CLR program objectives. A native of Minnesota, Ms. Blixrud received her B.A. from Augsburg College and the M.A. in library science (with related fieldwork in management and management information systems) from the University of Minnesota. ■■

OCLC Completes Installation of $70 Million Telecommunications Network

OCLC completed the installation of its $70 million telecommunications network on November 26, 1991. An X.25 packet-switched network that supports the Open Systems Interconnection standard, the New Network provides telecommunications service to nearly fourteen thousand OCLC-participating libraries in forty-six countries.

OCLC ran two telecommunications networks during the fourteen-month installation period. The original network was a dedicated, leased-line, 2400bps network with all circuits connected to OCLC in Dublin, Ohio. The new OCLC telecommunications network has forty-five nodes in forty-five cities across the country connected in an X.25 packet-switching net-

work, with the OCLC protocol in use from the node site to the end-user location. It is designed to operate at transmission speeds ranging from 2400bps to 1.5Mbps. Major backbone nodes in six cities—Atlanta, Chicago, Dallas, Los Angeles, New York, and Washington, D.C.—are connected with one another and OCLC in Dublin.

George Carpenter, director, Network and Hardware Services Division, and an OCLC staff of forty engineers and technicians worked with staffs of US Sprint and NCR to replace 3,367 modems linked to more than 10,000 workstations in OCLC member libraries across the entire United States. They converted approximately 60 modem sites each week. At the same time, libraries continued to use the OCLC system for cataloging and resource sharing at record levels.

According to Carpenter, the New Network is performing well. "Overall message traffic on the New Network is averaging about three million messages a day," he said. "Network availability is 99.7 percent, which is higher than the old network's availability of 99.6 percent."

"The distributed network architecture and standard X.25 protocol enable OCLC to link more easily with other networks in the education and scientific communities," said Fred Lauber, director, OCLC Systems and Telecommunications Engineering Division, and project manager for the New Network. He noted that OCLC had already implemented its first X.25 link with the internal network of the New York Public Library in June 1991.

Installation of the New Network began in November 1990 and was completed on November 26, 1991, a full month ahead of OCLC's planned completion date of December 31, 1991.

The New Network comprises 774 interstate and intrastate circuits—more than 280,000 miles of telephone lines. It also supports more than four thousand dial-access users, who access OCLC systems through either the New Network or CompuServe. At the same time it began to deploy the New Network, OCLC also began to install its new online system for cataloging and resource sharing, the PRISM service. Libraries could convert to the PRISM service after they had been transferred to the New Network. After twelve months, 80 percent of cataloging libraries have migrated to the PRISM services, and all OCLC cataloging institutions will be online to PRISM by April 1992. Users can also access other OCLC online products, including the EPIC service and the FirstSearch Catalog via the New Network. ■■

Rob Carlson Named Interim LITA Program Officer

Rob Carlson has been appointed interim program officer for the Library and Information Technology Association (LITA). He assumed his new position January 2, 1992. Carlson was manager of the American Library Association's Information Technology and Publishing and ALANET, the electronic-mail service. ALANET discontinued operation as of January 31, 1992.

"We are fortunate to obtain the services of Rob Carlson to fill this position for the next year," said Linda Knutson, LITA executive director. "Rob will primarily work on the planning of the LITA National Conference in September 1992."

Carlson replaces Nancy H. Evans, who will become head librarian at Pennsylvania State University, Ogontz campus, in Abington. "Nancy has contributed greatly to the development of this position," said Knutson. "I look forward to her continued contributions as an active LITA member." ■■

Exhibitors' Prospectus for LITA's September National Conference

The exhibitors' prospectus for LITA's Third National Conference is now being distributed. To be held in the new Colorado Convention Center in Denver, September 13–17, 1992, the conference's theme is "Information Technology: It's for Everyone."

More than 2,500 library and information professionals are expected to attend the LITA conference. The LITA conference, previously held in 1983 and 1988, provides the best opportunity to focus on automa-

tion and information technology for libraries.

"The key decision makers are among the wide variety of information professionals that will attend," said Betty Bengtson, chair of the national conference and director of the University of Washington Libraries. "Forty-five percent of the exhibit hours will be no-conflict, and according to past conference evaluations, ninety-nine percent of attendees viewed the exhibits."

Copies of the exhibitors' prospectus are available from the LITA office, (312) 280-4270, fax (312) 280-3257. For questions on exhibits, contact exhibits manager Sandy Donnelly, (708) 969-7988, fax (708) 969-8045. ■■

Conservation Education Programs at UT Austin

The Conservation Education Programs of the School of Library Service at Columbia University will be joining the Graduate School of Library and Information Science (GSLIS) at the University of Texas at Austin in July 1992.

Brooke E. Sheldon, dean of the GSLIS, stated: "We are absolutely elated that the Columbia University Conservation Education Programs, which have set the standards of excellence in this field, will find a new home at UT Austin. This extends and builds on the Graduate School of Library and Information Science's existing strengths in archives and preservation and is a perfect fit for the needs of the Harry Ransom Humanities Research Center (HRHRC), the Perry-Castaneda Library, and other major research libraries on campus. In assembling the resources to accommodate the new programs, we have had tremendous support from our administration, as well as from Harold Billings, director of the General Libraries, and Tom Staley, director of the HRHRC."

In noting the decision, Professor Carolyn Harris, director of the Conservation Education Programs, said: "The University of Texas at Austin provides a vital library school environment, a strong archives program, one of the largest library conservation programs within the HRHRC, a growing preservation program in the General Libraries, as well as a local community of conservators, librarians, and archivists."

The University of Texas at Austin is already known not only for the preservation administration courses offered in the Graduate School of Library and Information Science, but also for the conservation program in the Harry Ransom Humanities Research Center and the preservation program in the General Libraries. The HRHRC program is internationally recognized for its contributions to conservation treatment, education, and research, and the General Libraries established a preservation program in the early 1980s. UT Austin presently employs 22.25 FTE employees working in preservation and conservation—a staff one-third larger than the average for members of the Association of Research Libraries.

The programs, which will be named the Preservation and Conservation Education Programs for Libraries and Archives (PCEP), recognize the need to maintain an uninterrupted flow of trained professionals, even while integrating these programs into the GSLIS. Accordingly, the preservation administrator component will begin in September 1992, and the conservator program will begin in academic year 1993–94. While settling in its new home, PCEP will take the opportunity to reevaluate and strengthen the curriculum to better continue serving national needs in the preparation of individuals capable of becoming tomorrow's leaders in the challenge of preserving the knowledge and cultural heritage of our society.

The University of Texas at Austin, through its Graduate School of Library and Information Science, plans not only to continue but also, through a program of aggressive fundraising and expanded laboratory space, to provide a base from which PCEP can build on the tradition that it brings with it. The programs remain unique in the United States and internationally as the only graduate-level academic resource for the preparation of conservators and preservation administrators for libraries and archives and are the principal national source of trained personnel in an expanding job market.

Those interested in beginning the program for preservation administration in the

fall of 1992 are advised to contact the Graduate School of Library and Information Science at the University of Texas at Austin, 78712-1276; (512) 471-3821. ■■

NAL Hosts First International Technology Workshop

The National Agricultural Library (NAL) will host the first New Technology Workshop, sponsored by the International Association of Agricultural Information Specialists (IAALD), at the library in Beltsville, Maryland, on September 16–20, 1991.

IAALD is a seven-hundred-member body of agricultural information specialists representing more than eighty countries. It was established in 1955 with the objective of promoting and advancing public education and information transfer in agricultural development and production.

The workshop will feature in-depth lectures and demonstrations on new information-management technologies. Workshop organizers said there will be opportunities for hands-on use of the technologies with guidance by internationally recognized experts. The registration fee for the workshop, which will include four lunches, is $490.

"Taking full advantage of recent developments in computer technology is vitally important to information specialists everywhere," said NAL Director Joseph Howard, who was recently elected to a five-year term as IAALD president. "If it is used wisely, the technology will allow the rapid spread of knowledge in the agricultural sciences."

Among topics scheduled to be covered in the workshop are telecommunications and networking, multimedia CD-ROM development, scanning and optical-character recognition, hypertext and expert system applications, digital-image transmission, informatics (simulations and forecasting), and machine-readable storage and access for photographs.

For further information on the IAALD New Technology Workshop, contact Gary K. McCone, Information Systems Division, National Agricultural Library, 10301 Baltimore Blvd., Beltsville, MD 20705; (301) 344-3813. ■■

IRIS Project Chosen as Joe Wyatt Challenge Success Story

The University of Houston Libraries' Intelligent Reference Information System (IRIS) project has been chosen by EDUCOM's Educational Uses of Information Technology (EUIT) program as one of its Joe Wyatt Challenge Success Stories. The Joe Wyatt Challenge was intended to identify 100 successful applications of information technology in the United States and Canadian colleges and universities; 101 projects were actually chosen.

The two-year IRIS project installed a ten-workstation CD-ROM LAN that provided users with access to nineteen citation, full-text, graphic, and numeric databases. It also developed an expert system, Reference Expert, which recommends more than three hundred electronic and printed reference sources. The expert system, which is written in PDC Prolog, is available on network and stand-alone 80386SX IBM-compatible workstations. Because the system's knowledge base is contained in ASCII files, other institutions can modify the knowledge base for local use. The IRIS project also conducted three research studies: (1) a performance bench mark for the CD-ROM LAN, (2) a study of user perceptions of the CD-ROM LAN, and (3) a study of user perceptions of the expert system.

The IRIS project was partially funded (approximately 51 percent) by a $99,852 Research and Demonstration Grant from the U.S. Department of Education's College Library Technology and Cooperation Grants program.

Reference Expert is available at no charge. A sixteen MHz 80386SX computer with an EGA or a VGA monitor, 1 MB of RAM, and a hard disk is the minimum recommended hardware configuration. To obtain a copy of the expert system, send a stamped, self-addressed diskette mailer and a 5 1/4-inch 1.2 MB or a 3 1/2-inch 1.44 MB diskette to Charles W. Bailey, Jr., Assistant Director for Systems, University Libraries, University of Houston, Houston, TX 77204-2091. Requests not accompanied by a mailer and diskette cannot be filled.

For further information, contact Charles W. Bailey, Jr., (713) 749-4241 or LIB3@UHUPVM1 (BITNET). ■■

Recent Publications

Book Reviews

Advances in Library Resource Sharing, V.1. Ed. by Jennifer Cargill and Diane J. Graves. Westport, Conn.: Meckler, 1990. 238p. $55 (ISBN 0-88736-490-X, ISSN 1052-262X).

Librarians have expended immense effort and financial resources in the last century on cooperative efforts to improve information access for their patrons. These resource-sharing projects have included bibliographic control, interlibrary loan, cooperative collection development, reciprocal borrowing, union catalogs, and centralized storage schemes. Rapidly increasing serials costs, the diversity of standards in new technology, and accreditation standards that emphasize collection size over access indicate that goals of improved access will not be easily achieved in the future. In this first volume of the *Advances in Library Resource Sharing* series, editors Cargill and Graves have chosen a fine collection of essays to illuminate the history, current practices, and future trends of resource sharing.

Richard M. Dougherty and Carol Hughes' historical survey details the trends and programs that have attempted "to promote the acquisition, control, and mobility of materials among libraries." The essay covers established programs such as CRL, OCLC, and RLG, as well as less well known but conceptually innovative projects such as the National Periodicals Center.

Among the four essays on traditional means of resource sharing is Michael Carpenter's "How Can We Improve Resource Sharing? A Scholar's View," which should be read by all reference and interlibrary loan personnel. The view from the other side of the desk proves dismaying! The suggestion for improvement by utilization of a library citation index system is a thought-provoking alternative that might motivate reluctant libraries to provide access to their unique materials.

The five essays that examine current resource-sharing practices are important because of the variety of local, regional, and multitype libraries discussed. Pamela Zager's paper covers five CD-ROM projects in Texas and illustrates the powerful potential of CD-ROM union catalogs in local or regional multitype library settings. Carolyn Grangaard Smith's essay on the Illinois Library and Information Network (ILLINET) recounts the success of the ILLINET fax network and the importance of state library funding. Suzanne Fedunok's "METRO Collection Inventory Project: A Conspectus Case Study" provides a review of earlier collection-assessment efforts and the evolution of METRO's collection-inventory model.

The future of resource sharing is no less challenging. Technological advances that minimize distance between patrons and desired information are also creating new dilemmas. A recurring, ominous theme suggests that libraries are now at a decisive point in determining their involvement in providing future information services. Strong and responsive competition is already being offered by commercial information vendors. Increased electronic publishing raises problematic questions about copyright and content control of a work. Among the four essays on the future, Bonnie Juergens and Gloriana St. Clair's essay offers a model for author accountability and copyright compliance in an electronic environment.

Each essay includes a bibliography. The volume concludes with an annotated list of suggested readings. An index would have been useful for locating particular projects throughout the essays.

This initial volume lays a groundwork for understanding past and present resource-sharing efforts, with some attempt to see into the future. Upcoming volumes that provide information about the trends of resource sharing, whether national or regional in scope, will be an asset to any library attempting to enhance services to their users through resource sharing.—*Kathleen O'Connor, Gonzaga University, Spokane, Washington.* ■■

Chen, Ching-chih. *Optical Discs in Libraries: Use & Trends.* Medford, N.J.: Learned Information, 1991. 237p. paper, $79.50 (ISBN 0-938734-49-0).

Chen has amassed more than 230 pages of statistics, graphs, and bibliography on the use of CD-ROMs in libraries during the period 1987–1989, but nowhere in the text is the purpose of these figures, charts, and lists clearly stated. The author comes closest to making a statement of purpose with these two sentences from the introduction: "Yet, how are these (CD-ROM) products and services used? Is CD-ROM a panacea?" (p.3). Both queries, hopelessly vague, beg such further questions as the following: How will a study of CD-ROM use be beneficial? To whom will it be beneficial? Are there any real questions that information professionals have concerning the use of this storage technology? Are there issues and questions in the literature that have gone unanswered? Is this study going to answer these questions? Is CD-ROM a panacea for what? The posing and answering of these and other questions in the introductory chapter might have given the work focus and direction. As it stands, the book is a mere collection of data.

The book's first section consists of tabular and graphical displays of data collected from Chen's 1987–88 study of CD-ROM use in North America and from a 1989 study of CD-ROM use in Europe conducted by Chen and David Raitt. The survey questionnaire used in the study asked libraries about their use of CD-ROM products, the effect CD-ROMs have had on use of print and online sources, the number of CD-ROM players they have, and so on. The point of these queries is unclear, and the statistics resulting from the studies are inevitably outdated in a field that changes almost weekly.

The second part of the work contains a bibliography, with abstracts, of the literature through 1990 on CD-ROM use in libraries. The abstracts derive from the cited articles themselves or are supplied by the author. An index to this bibliography gives access by product and then by the subheadings "Treatment," "Institution/Location," and "Author(s)/ Main Entry." The majority of the articles listed are "experience" reports that were perhaps quite helpful to librarians during the years 1986–90.

Beyond the uncertain value of the contents of Chen's book, the presentation, grammatical and otherwise, is unquestionably abominable. After reading a few pages, one wonders if anyone proofread the manuscript before it went to the publisher. Even more disturbing, one must assume that the publisher did not proofread and edit the text before it went to press. One or two examples of these unfortunate errors will suffice:

"Nevertheless, since four fifth (*sic*) of the total number of CD-ROM products have originated in the North America (*sic*), it is . . ." (p.76).

"Figure 37 shows the comparative responses on the impacts of optical products as found American and European (*sic*) studies" (p.78).

With the presence of so many flaws in presentation and with the absence of purpose or timeliness in content, *Optical Discs in Libraries* has little to recommend it. Any benefit derivable from the book's extensive bibliography and indexes pales in the face of the book's $79.50 cover price.—*Brian Sealy, University Library, University of Michigan, Ann Arbor.* ■■

Dempsey, Lorcan. *Libraries, Networks and OSI: A Review, with a Report on North American Developments.* Bath, U.K.: U.K. Office for Library Networking, The Library, University of Bath, 1991. 232p. paper, $60 (ISBN 0-9516856-0-0).

There is a growing perception that computer networks will play a pivotal role in the delivery of library services in the 1990s. Leaving aside library resource-sharing

"networks" that are not inherently computer based, libraries are already involved in providing services on institutional local area networks, specialized state and regional networks (e.g., ILLINET Online), and national and international computer networks (e.g., the Internet). They have played a major role in developing and supporting the bibliographic utilities, which are increasingly providing end-user services. Open Systems Interconnection (OSI), a complex group of standards organized into a seven-layer model, promises to help bind together these diverse networking efforts, but OSI is still evolving.

In the form of a report to the British Library's Research and Development Department, Lorcan Dempsey has given us a valuable discussion of important aspects of OSI plus an in-depth look at state-of-the-art networking projects in the United States and Canada. He also provides a tantalizing peek at United Kingdom activities in brief commentaries at the end of each chapter. The international perspective of this book is appealing. Global computer networks will shrink the world more than previous technologies, and they will offer new opportunities for cooperative library services as well as new challenges. We need similar books dealing with library-oriented networking activities in other parts of the world.

After an introductory chapter, the next five chapters of the book provide detailed discussions of OSI-related topics, including the following: OSI protocol model; ASN.1; the difference between interconnection, interoperability, and interworking; X.400; EDI; X.500; FTAM; the Linked Systems Project; the ISO Interlibrary Loan Protocol; the ISO Search and Retrieve Protocol; and Z39.50. Dempsey often follows a theoretical discussion of an OSI standard with an interesting summary of projects that employ that standard. He carefully analyzes the problems inherent in utilizing different OSI standards. I particularly liked his discussion of the steps that need to follow the establishment of a standard: profiles, conformance testing, and registration. By understanding these processes, we can develop more realistic expectations about how long it will be after a standard is established until interoperable systems that employ that standard are available.

The reader unfamiliar with OSI concepts will find this part of the book slow going; however, Dempsey writes clearly, and he usually defines acronyms and terms as they occur. Since the combination of OSI and libraries as topics guarantees alphabet soup, the acronym and standards lists at the back of the book are helpful. Overall, this section of the book is dense but rewarding.

The remaining three chapters go more quickly. In them, Dempsey looks at libraries and research networks (e.g., BITNET), local public-access computer systems, and resource-sharing networks. He discusses a number of topics, including the following: major North American and UK research networks, network publishing, Carnegie Mellon University's Mercury Electronic Library Project, the electronic publishing efforts of the Welch Library at Johns Hopkins University, the Corporation for National Research Initiative's Digital Library project, the Coalition for Networked Information, local electronic library resources, some individual library systems, the bibliographic utilities, ILLINET Online, OLIS (now OhioLINK), MELVYL, CARL, and the Irving Library Network. If you have been closely following the journal literature and the library-oriented list servers, much of this material will seem familiar; however, Dempsey has done a good job of organizing and presenting it, and it is useful and instructive to have it gathered together in one place.

Throughout the book, it is clear that Dempsey has done his homework. He benefited from a rapid-fire study trip to sites of major North American library networking activity, and he utilized list servers and electronic publications in addition to more conventional sources. The book contains an extensive bibliography.

This book is highly recommended for librarians who wish to obtain more in-depth understanding of computer networks and their implications for libraries. If you are completely new to OSI, you might want to read a general overview article, such as Ray Denenberg's "Data Communications and OSI" article in the fourth issue of the 1990 volume of *Library Hi Tech*, prior to tackling this book. [Ed. note: Since completion of this review, a U.S. edition has been published: Dempsey, Lorcan. *Libraries, Net-*

works, and OSI: A Review with a Report on North American Developments. Westport, Conn.: Meckler, 1992. 232p. paper, $49.50 (0-88736-818-2).]—*Reviewed by Charles W. Bailey, Jr., Assistant Director for Systems, University Libraries, University of Houston.* ■■

Kim, David U., and Douglas M. Kim. *Policies of Education Software Publishers: A Guide for Authors.* The Woodlands, Tex.: New Technology Pr. (P.O. Box 9154, The Woodlands, TX 77387), 1991. 231p. paper, $25.

This publication is a concise reference tool, meant to be used by authors of educational software who are seeking prospective publishers. The bulk of the book (194 pages) is an alphabetic listing of nearly two hundred publishers of educational software, giving basic information regarding submission of software, review, publication, and payment policies. Four additional indexes to publisher names listings appear, including subjects, product names, and computer systems.

The information contained in the work was gathered from questionnaires sent to the publishers. Some of the descriptions are quite brief, stating only that the publishers do not seek outside software. Most descriptions are more than a page long and provide useful information regarding the publisher's interests and general policies governing submission and acceptance of software produced by independent developers.

For many of the software publishers the guide includes additional information, such as promotional effort offered, lists of titles in print, and lists of top-selling products. This information provides insight into the size and scope of their educational software operations.

A minor criticism of this work would be its plain appearance. It contains no graphics or variations in type, and lacks right-justified margins. For a work on educational software, it seems odd that the authors did not make use of current desktop publishing capabilities.

This criticism aside, the authors have compiled a compact, fact-filled publication that they plan to review and update regularly. For the growing number of classroom teachers, college faculty, and others who are producing educational software and wish to have it published, the information provided by this work could be very helpful.—*Charles Litchfield, Virginia Polytechnic Institute and State University.* ■■

Mount, Ellis, and Betrice Kovacs. *Using Science and Technology Information Sources.* Phoenix, Ariz.: Oryx, 1991. 189p. $26.95 (ISBN 0-89774-593-0).

This book would have been more appropriately entitled "An Introductory Survey of Information Formats" because it provides an extensive listing with descriptions for every conceivable type of information format, each one covered in a separate chapter, listed alphabetically under three major categories: primary sources of information, secondary sources of information (textual), and secondary sources of information (nontextual). Despite the apparent comprehensiveness of this approach, one is mystified to find no separate chapter covering abstracting and indexing services. In fact, some very brief descriptive material (three paragraphs only) is included in the chapter "Journals and Periodicals" in the section on primary sources of information. In keeping with this obscuring of the crucial roles played by abstracting and indexing services, the authors omit them entirely from the glossary of terms appended to the text. Puzzling indeed.

The chapter "Computerized Information Sources" is another disappointment. The basic descriptions of the functions and applications of online database systems and CD-ROM products are lacking in clarity and are poorly organized. Separate chapters on CD-ROM and online products would have been preferable as a minimal approach to appropriate coverage of these important topics. Despite the fact that numerous references are made to CD-ROM and online systems in various parts of the text, once again these terms are not included in the glossary. Compounding these problems are errors, such as the assertion that *Chemical Abstracts* is available in CD-ROM format (p.29); it is not, nor is CASSI (*Chemical Abstracts Service Source Index*) available either online (p.32) or on CD-ROM. However, in an error of omission, no mention is made of the MEDLINE

CD-ROM product (or numerous others that might have been noted).

The authors declare in the preface that "[t]his book is *not* a bibliography of sci-tech sources" (preface, p.vi), and this is certainly true enough. However, it is a further weakness that too few examples of specific sources were included in the discussions of each format to provide the reader with a clear picture of the diversity and richness of the materials available, let alone explain how to use them as implied by the title of the work: typically no more than a half dozen sources per chapter are included, and the descriptive coverage is very brief. In summary, this book cannot be recommended for any of the three purposes outlined by the authors in the preface. It would not serve as a satisfactory textbook for library students, nor would it have much value for experienced sci-tech librarians. Finally, it is difficult to see how this book could be used "to assist librarians with collection management of sci-tech literature" (preface, p.v).—*Howard M. Dess, Library of Science and Medicine, Rutgers University, New Brunswick, New Jersey.* ■■

Orion Blue Book: Computer 1991. Durango, Colo.: Orion Research Corporation (1315 Main Ave., Ste. 230, Durango, CO 81301), 1991. 544p. $129 (ISBN 0-93289-50-X).

Orion Blue Book: Computer 1991 is a price list of used personal computer equipment and peripherals. It also includes some mini- and midrange computer systems such as the AT&T 3B2 series, the HP 3000, and the IBM AS/400. Compiled annually from dealer surveys, the *Blue Book* lists dealers' average resale prices for used products as well as the average prices paid to customers for equipment in both mint and standard conditions.

The book is organized alphabetically by manufacturer and for every listed product includes a description of each item; the year of manufacture; the model name, number, and configuration; the new list price; the retail used price, i.e., the average price for which the product can be sold within thirty days; the wholesale mint price, i.e., the average price paid to a customer for the product in excellent condition and working order; and the wholesale average price, i.e., the average price paid to a customer for equipment in good, rather than mint, condition.

Because the book has been compiled from surveys of computer dealers who have chosen to participate, rather than from surveys of all dealers in the U.S., not all equipment may be listed. If a particular item is not listed, it means only that it has not come onto the market with any of the participating dealers. It does not mean that it has no market value or that it cannot be bought or sold on the used equipment market. Also, since the prices listed are the average prices nationwide, actual prices may vary slightly from region to region.

The listings are far more extensive and include more manufacturers and models than the *NACD Computer Blue Book: The Official New and Used Computer Price Guide* published by the National Association of Computer Dealers. The *Orion Blue Book*, however, does not include the address directory of used computer dealers, manufacturers, service companies, and publishers provided in the NACD book.

The *Orion Blue Book* would be a useful reference work for any organization—whether business, university, or large library system—that owns, buys, sells, or insures large quantities of personal computer equipment. Because of its potential value to a wide segment of the public, libraries might consider adding *Orion Blue Book: Computer 1991* and successive editions to their general, business, or computer reference collections. With the continued growth in personal and midrange computing, the *Orion Blue Book* has the potential to become as valuable a reference resource as the automotive blue books, which are now standard references for information on used car prices.—*Kathleen Rehn, The New York Public Library.* ■■

The Reference Library User: Problems and Solutions. Ed. by Bill Katz. New York: Haworth, 1991. 151p. $24.95 (ISBN 1-56024-022-90).

A collection of eleven papers addressing various aspects of reference service, this latest editorial accomplishment by Bill Katz will be of interest to all types of libraries. Of particular benefit is the chapter on special populations; contributions here include

treatments of deinstitutionalized (those who have been released from mental or other hospitals with emotional, physical, or mental disorders), older, learning-disabled, blind, and physically handicapped readers. The references to other works dealing with these populations are especially helpful to those wishing to do further research. Sheila Intner's offering on missed opportunities for bibliographic instruction is again relative to all users and nonusers in all types of libraries. She states that we need to ask ourselves "Who is our public?" and "What are we teaching them?" Very helpful to any library staff member is Rubin's "Anger in the Library: Defusing Angry Patrons at the Reference Desk (and Elsewhere)," because everyone who works with the public will encounter this behavior at some point. Her explanation for why these encounters are on the rise is especially interesting. Buschman's discussion of information brokers compares their services to a traditional reference model. The possibilities she poses in her conclusion as to the direction the profession is taking are thought-provoking. Cannon's look at academic library nonusers provides an interesting view of student and faculty lifestyles in an attempt to explain how they use the library or why they don't. A notable statement: "Perhaps those students most likely to be nonusers are those never exposed to formal library instruction or to instruction for the particular academic library they are now expected to use" (p.126). What does this say to academic librarians, faculty, and administrators? It bears examination.

While this is not a mandatory purchase, it is certainly a good one for the price. There are some valuable information and tangible suggestions here that will help us as a profession look critically at our users and decide how they are best served.—*Lisa K. Miller, American Graduate School of International Management Library, Glendale, Arizona.* ■■

Other Recent Receipts

Advances in Classification Research: Proceedings of the 1st ASIS SIG/CR Classification Research Workshop. Ed. by Susanne M. Humphrey and Barbara H. Kwasnik. Medford, N.J.: Learned Information, 1991. 172p. paper, $39.50 (ISBN 0-938734-53-9).

Budgets for Acquisitions. Ed. by Sul Lee. New York: Haworth, 1991. 134p. $22.95 (ISBN 1-56024-158-6).

Cataloging: The Professional Development Cycle. Ed. by Shelia S. Intner and Janet Swan Hill. Westport, Conn.: Greenwood, 1991. 159p. $37.95 (ISBN: 0-313-27254-9).

Cone, Robert J. *How the New Technology Works: A Guide to High-Tech Concepts.* Phoenix, Ariz.: Oryx, 1991. 123p. paper, 28.50 (ISBN 0-89774-652-X).

Creative Planning for Library Administration: Leadership in the Future. Ed. by Kent Hendrickson. New York: Haworth, 1991. 113p. $19.95 (ISBN 1-56024-093-8).

Davies, Peter. *Artificial Intelligence: Its Role in the Information Industry.* Medford, N.J.: Learned Information, 1991. 114p. paper, $39.50 (ISBN 0-938734-50-4).

Describing Archival Materials: The Use of the MARC AMC Format. Ed. by Richard P. Smiraglia. New York: Haworth, 1991. 228p. $29.95 (ISBN 0-086656-916-2).

Ensor, Pat. *CD-ROM Research Collections: An Evaluative Guide to Bibliographic and Full-Text CD-Rom Databases.* Westport, Conn.: Meckler, 1991. 302p. $55 (ISBN 0-88736-779-8).

Essential Guide to Multifunction Optical Storage. Ed. by Judith Paris Roth. Westport, Conn.: Meckler, 1991. 134p. $40 (ISBN 0-88736-751-8).

The Evolution of Library Automation: Management Issues and Future Perspectives. Ed. by Gary M. Pitkin. Westport, Conn.: Meckler, 1991. 190p. $42.50 (ISBN 0-88736-811-5).

INSTRUCTIONS TO AUTHORS

Information Technology and Libraries welcomes manuscripts related to all aspects of library information technology. Some specific topics of interest are mentioned on the masthead page. Feature articles, communications, letters to the editor, and news items are all considered for inclusion in the journal. Feature articles are refereed; other items generally are not. All material is edited as necessary for clarity and length.

Manuscripts must be typewritten and the original submitted with one duplicate. All text must be double-spaced, *including footnotes and references*. Manuscripts should conform to *The Chicago Manual of Style*, 13th ed., rev. (Chicago: Univ. of Chicago Pr., 1982). Illustrations should be prepared carefully as camera-ready copy, neatly drawn in a professional manner on separate sheets of paper. Manuscript pages, bibliographic references, tables, and figures should all be numbered consecutively. Final versions of papers accepted for publication must be submitted on floppy disk (5.25" or 3.5") in *WordPerfect* (DOS versions 4.2, 5.0, and 5.1 accepted) along with two printed copies. If *WordPerfect* is not available, an ASCII copy is acceptable.

Feature Articles consist of original research, state-of-the-art reviews, or comprehensive and in-depth analyses. An abstract of one hundred words or less should accompany the article on a separate sheet. Headings should be used to identify major sections. Authors are encouraged to relate their work to other research in the field and to the larger context of economic, organizational, or management issues surrounding the development, implementation, and use of particular technologies.

Communications consist of brief research reports, technical findings, and application notes. An abstract need not be included.

Letters to the Editor may offer corrections, clarifications, and additions to previously published material or may be independent expressions of opinion or fact related to current matters of concern in the interest area of the journal. A letter commenting on an article in the journal is shared with the author, and a response from the author may appear with the letter.

News and Announcement items may announce publications, conferences, meetings, products, services, or other items of note.

Book Reviews are assigned by the book review editor. Readers wishing to review books for the journal are invited to contact the book review editor, indicating their special areas of interest and expertise.

Software Reviews are assigned by the software review editor. Readers wishing to review library-related software for the journal are invited to contact the software review editor, indicating their special areas of interest and expertise.

Names and addresses of the journal editors may be found in paragraph two on the masthead page. In all correspondence please include your own name, institutional affiliation, mailing address, and phone number.

SETTING STANDARDS IS EVERYONE'S BUSINESS.

Establishing effective technical standards. It makes good business sense, no matter what business you're in. Because clear, cohesive practices and procedures help things run smoothly.

Which is why we need your help.

Who are we? We're NISO. The National Information Standards Organization. We develop technical standards for the entire range of information services and products: online information services, librarianship, bookselling, indexing, and library equipment and supplies. In fact, our standards are an integral part of many work environments.

NISO has developed over fifty information standards. Here are just a few of the standards that benefit you.

- Standard identifiers for books and serials
- Standard format and vocabulary for information sharing
- Standard computer-to-computer protocol linking local, regional, national, and international bibliographic systems
- Standards for permanent and durable book papers and bindings

How can you help us? By helping NISO set the standards in your profession. Because only with the input of successful organizations like yours can we develop proper standards. We also need your help with funding. You see, we depend on the support of our members. Members from all types of organizations including libraries, publishers, large and small companies, and government agencies. There are a range of membership options available; one is right for your organization.

What's in it for you? A chance to influence your profession's direction and growth by setting its standards. A chance to work closely with your peers. A chance to plan for the new standards that will directly affect you.

What should you do now? Call, write, or fax us. We'll tell you more about ourselves, how we can work for you, and how you can help us.

NISO

P.O. Box 1056 • Bethesda, MD 20827
Telephone: 301-975-2814 • Fax: 301-869-8071 • Bitnet: NISO@NBSENH