Volume 5
Number 2

ISSN 0730-9295

Information Technology and Libraries

June 1986

CONTENTS

ITLBDC 5(2) 89–176 (1986); Inf Technol Libr

Experience counts in authority control. And if your library is choosing a vendor to apply authority control to your MARC records before you put them online, consider the following:

Blackwell North America has 12 years experience performing retrospective automated authority control edits.

B/NA has performed subject authority control on hundreds of MARC databases since LC began issuing machine readable subject authority data in 1973; and name authority control on over 150 MARC databases since LC began issuing machine readable name authority data in 1982. In addition, B/NA was first to offer library specific authority files for loading into online systems. B/NA's staff of authority control editors has combined experience of over 33 years, and applies manual corrections to your bibliographic records where the automated system stops.

B/NA is interested in your project, and will deliver professional service on schedule.

★ B/NA AUTHORITY CONTROL CUSTOMERS

University of Illinois at Chicago
University of Wisconsin, Eau Claire
Grove City Public Library
Vanderbilt University
University of Notre Dame
Ball State University
College of William & Mary
University of Texas, El Paso
Warren Trumbull County Public Library
University of Oregon
Lamar University
University of Delaware
Rollins College
Louisiana State University
Johns Hopkins University

For more information on authority control for your library, contact B/NA's authorities on authority control at the address below.

BNA
BLACKWELL NORTH AMERICA, INC.
6024 S.W. Jean Road, Building G
Lake Oswego, OR 97034
Phone toll-free 800-547-6426

Information Technology and Libraries

Volume 5, Number 2: June 1986

CONTENTS

ITLBDC 5(2) 89–176 (1986); Inf Technol Libr

Information Technology and Libraries

Information Technology and Libraries 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:* Donald P. Hammer. 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 80 of the March 1986 issue.

Manuscripts of articles, communications, and news items should be addressed to: William Gray Potter, Editor, *Information Technology and Libraries*, Hayden Library, Arizona State University, Tempe, Arizona 85281. Copies of materials for review should be addressed to: Karin A. Trainer, *ITAL Book Reviews*, Yale University Library, 120 High St., P.O. Box 1603A, Yale Station, New Haven, CT 06520. Advertising arrangements should be made with William Z. Schenck, University of Oregon Library, Eugene, OR 97403.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $12.50, is included in membership dues. Nonmembers may subscribe for $25 per year. Single copies, $7.50.

Circulation and Production: American Library Association, 50 E. Huron St., Chicago, IL 60611. Please allow six weeks for change of address.

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 *Computer Contents, Computer Literature Index, Current Contents, Current Index to Journals in Education, Education, Library Literature, Magazine Index, and NewSearch*. Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

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.*

Two new sources for Information Professionals

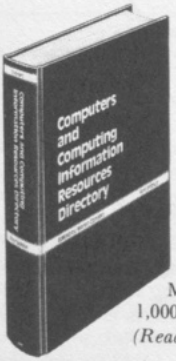
Computers and Computing Information Resources Directory

This new work is a one-stop, annotated, and indexed guide to the wide array of organizational and published sources of information on the technical, economic, and social implications of computers and information processing.

Over 6,000 entries cover the primary sources of information in the computer world, including associations and user groups, consultants, training organizations, special libraries, information centers, research centers, university computer facilities, computer trade shows, online services, teleprocessing networks, and much more.

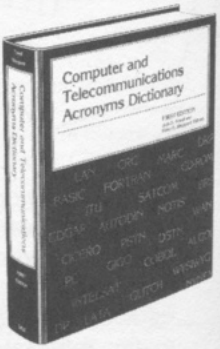
Each entry contains organization name or publication title, address, telephone number, key executive names, and, in most cases, a complete description.

Easy access to the entries is provided by the Master Index, the Geographic Index, and the Personal Name Index.



First Edition.
Edited by
Martin Connors.
1,000 pages. $160.00.
(Ready August 1986)

Coming this fall . . . from Gale



First edition.
Edited by Julie E. Towell
and Helen E. Sheppard.
About 1,200 pages. $60.00.
(Ready September 1986)

Computer and Telecommunications Acronyms Dictionary

Today's literature on computers and telecommunications is littered with acronyms like PSTN, ASCII, LANs, and NYNEX. The new **Computer and Telecommunications Acronyms Dictionary** fills the need for a comprehensive source for understanding these terms and using them correctly.

The dictionary contains over 25,000 acronyms taken from computer and telecommunications terminology and includes data processing terms, names of associations and periodicals, and information and telecommunications services. Each entry includes the acronym, its definition, and in many cases additional information such as subject category, geographic location, and citation of source.

Entries in the dictionary have been selected from Gale's **Acronyms, Initialisms, and Abbreviations Dictionary**, and also contains new entries not yet published in *AIAD*.

For Fast Service—
Order Tollfree: 800-223-GALE

All books are sent on 60-day approval. Available at Gale's 5% Standing Order Discount. Prepay and receive an additional 5% discount. Customers outside the U.S. and Canada add 10%.

Gale Research Company
Book Tower • Detroit, MI 48226

Cocited Author Retrieval

Howard D. White

Online searchers are introduced to retrieval on pairs of cocited authors, which is a useful complement to standard subject retrieval and may make Scisearch and Social Scisearch more accessible. The paired authors, read as indicative of subjects, generally produce hits not found by use of subject headings. Three cocited author search strategies are demonstrated: single pairs, systematic pairing of a small group of authors, and free pairing of many related authors from two lists. The last strategy was used innovatively to produce a mailing list of contributors to a transdisciplinary field, indicators of science and technology.

INTRODUCTION

Popular belief has it—probably correctly—that most online searchers would rather base searches on subject headings or their natural language counterparts than on cited references. Cited reference (CR) retrieval is of course identified with Scisearch and Social Scisearch (respectively, the online versions of Science Citation Index and Social Science Citation Index). Its basic form is widely familiar. Instead of searching on one or more nouns or noun phrases indicating a subject, one searches on a work that later authors may have cited and retrieves the papers that in fact cited it. Presumably these "descendents" are related to it in subject matter or methodology. Some find this kind of retrieval onerous because the cited work must be known in advance of going online. Further, it must be designated without error in a form that requires author's name, year of publication, and volume and page numbers; for example:

SELECT CR = GILBERT GN, 1977, V7, P113

It is unfortunate that cited reference retrieval of this sort puts some people off, because CR retrieval tends to complement, rather than duplicate, retrieval by subject headings on a given topic. In other words, it usually produces unique hits that subject retrieval, for a variety of reasons, fails to yield.

The object of this paper is to popularize an alternative technique for CR searching that reliably produces good results in many subject areas: cocited author retrieval. First introduced in *the Journal of the American Society for Information Science*[1] and discussed in several later articles,[2-5] it still is not common in the searcher's bag of tricks. Below, it will be illustrated on three levels: cocited authors formed in single pairs, systematic pairing of a small group of related authors, and free pairing of many related authors from two separate lists. The last was used in an interesting large-scale retrieval that is reported here for the first time.

The practical attraction of cocited author retrieval is that, in an online system, it can be undertaken immediately, with no more needed as search terms than what the searcher knows or can easily learn: the names of at least two key authors in a subject area or specialty. One does not need to know by name exactly what they have writ-

Howard D. White is associate professor, College of Information Studies, Drexel University, Philadelphia.

ten, or to have specific bibliographic details like those in the Gilbert example above. Nor does one need to know, or look up, subject headings, or to guess at various natural language expressions of the topic of interest (unless one wants to do a complementary subject search).

Undeniably, however, successful cocited author retrieval lies in choosing good names on which to search. They must be authors prominent enough to have been cited by other writers in journals recorded in Scisearch or Social Scisearch. The authors must also be related enough in other writers' eyes to be cited together (cocited) with some frequency. Lastly, the searcher or customer must know that certain names together imply a topic, which is to say that someone must be able to interpret and propose paired authors' names as if they were subject headings.

RETRIEVAL WITH SINGLE PAIRS OF AUTHORS

As stated in the *Journal* article, "Cocited author searches are performed by specifying two authors and retrieving any paper that cites both."[6] Take this example, using DIALOG conventions, of a potential retrieval in Social Scisearch:

SELECT CR = ARENDT H? AND
CR = NOLTE E?

Here the searcher is asking for all papers that cite anything by philosopher Hannah Arendt and historian Ernst Nolte. The truncator strips off the bibliographic identification of particular cited works, so that one obtains papers that cite any work by either Arendt or Nolte as long as both are included in the references. (N.B.: Cocited authors are not the same as coauthors. People who have written together are sometimes cocited, but coauthorship is not necessary for the technique to work.) The particular combination of Arendt and Nolte is viable: they have been cocited frequently in the journal literature because (although never coauthors) both have written books on or relevant to Nazism, fascism, and the Holocaust. Someone who can "read" this subject into their conjoined (ANDed) names will in fact retrieve many papers on it by putting the above statement into Social Scisearch.

It is worth noting that the statement

SELECT CR = ARENDT H?

by itself would retrieve all papers that cite anything by Hannah Arendt, who wrote on a great variety of topics. The citing papers would be comparably varied in subject matter. It is by constraining Arendt as a cited author with Nolte that we obtain a relatively well focused search. What may seem a coarse meshed style of retrieval is, in practice, not: ANDing well chosen authors discriminates nicely as a rule, even when one or both of them have written on many topics in the course of producing their oeuvres. (A few writers are simply too rich in their cultural connotations to work well; retrieval on a pair like Marx and Freud produces a farrago.)

Cocited author retrieval is actually retrieval based on the intersection of two oeuvres. Citing papers pile up on relatively few such intersections, but when they do, it generally means that both the intersecting oeuvres and the citing papers share a subject. Often, too, the subject linkages are more subtle than those that subject indexers ordinarily discern. In return for the inevitable false drops, one almost always finds some nice surprises—relevant papers whose existence would not be revealed by subject indexing of the usual kind.

In Scisearch and Social Scisearch, one can search only on surnames and initials and only on first authors in multiple authored works. These limitations affect cocited author retrieval in those databases. However, the technique has an interesting property in that it can usually distinguish authors whose names are homonyms. If, for example, one enters in Social Scisearch

SELECT CR = WHITE HD?

one retrieves papers that have cited works by the present author, but also by other H. D. Whites, including one who writes on Dutch economic history. However, if one ANDs my name with someone else's in library and information science, thus:

SELECT CR = WHITE HD?
AND CR = SMALL HG?

the other H. D. Whites drop out as bases of retrieval because no one cites them with Henry G. Small (a founder of cocitation analysis). Similarly, if one paired

CR = WHITE HD? with another cited author in economic history, retrievals would not be based on any papers of mine. Homonym discrimination breaks down only in those rare cases when two authors with identical surnames and initials publish in the same field (for example, CR = WILSON P? retrieves papers citing either Patrick Wilson or Pauline Wilson in library and information science, even when ANDed with another author in that field).

Figure 1 lists nine pairs of cocited authors and my labels for the topics they represent. They are known to produce hits when introduced into Social Scisearch. Those wanting to try cocited author searching might begin with one or more of these pairs and examine the results. It will be seen that, in some cases, cocited author pairs express topics that are not captured by any existing subject indexing.

The same technique works very well in Scisearch. For example, a search on "glucose transport and brush-border membrane vesicles" can be expressed as

SELECT CR = HOPFER U?
AND CR = ARONSON PS?

COMBINATION OF ALL POSSIBLE PAIRS

If the searcher or customer can name a small set of cited authors who are key contributors to a field, good retrievals can often be made by using all possible pairs of authors' names as search terms. *N* authors produce N(N – 1)/2 pairs. Six authors produce 6(5)/2 = 15 pairs. Obviously, the more authors one starts with, the more pairs one must enter (if all combinations are to be formed), and the possible combinations multiply quickly. Thus, for practical reasons, it is best to keep the input set of authors relatively small.

The example here uses DIALOG conventions with an input set of four social psychologists specializing in human judgment and decision making: Fritz Heider, Edward E. Jones, Harold Kelley, and Irving Janis. Jointly they connote attribution theory. ("Attribution Theory seeks to describe the processes by which people arrive at causal explanations . . . for behavior. A primary concern is with how people distinguish between personal and impersonal causes. . . ."[7] The source of the quotation mentions Heider, Jones, and Kelley as key figures.) One can retrieve a large literature on this topic in PsycInfo (almost four thousand items in August 1984) simply by using the descriptor Attribution. However, many additional (nonduplicating) items can be retrieved in Social Scisearch with the strategy:

SELECT CR = HEIDER F?
SELECT CR = JONES EE?
SELECT CR = KELLEY H?
SELECT CR = JANIS I?

Then, assuming these create sets 1, 2, 3, and 4:

Pairs	Topics
S CR = EPSTEIN C? AND CR = ROSSI A?	Occupational sociology of women
S CR = BERLYNE D? AND CR = MONTGOMERY K?	Curiosity and exploration in animals
S CR = CRANE D? AND CR = MULLINS N?	Communication patterns in science
S CR = HAGERSTRAND T? AND CR = ROGERS E?	Geographical diffusion of innovations
S CR = JENCKS C? AND CR = JENSEN A?	Race, IQ, and schooling
S CR = KEY VO? AND CR = BURNHAM W?	Critical elections and voting behavior
S CR = STEWARD J? AND CR = THOMAS DH?	Prehistoric Great Basin ecology
S CR = CUTRIGHT P? AND CR = AARON H?	National welfare programs
S (CR = ATHERTON P? OR CR = COCHRANE P?)* AND CR = BATES M?	Online search strategies

*Pauline Atherton Cochrane's name is given in both forms in which it may be cited.

Fig. 1. Sample Cocited Author Pairs and Their Implied Topics.

COMBINE 1 AND (2 OR 3 OR 4)
COMBINE 2 AND (3 OR 4)
COMBINE 3 AND 4

Assuming these create sets 5, 6, and 7:
COMBINE 5-7/OR

In August 1984 the last combination retrieved a literature of more than 1,300 items, of which about 450 have some form of the word *attribution* in their titles. (Items retrieved by more than one combination of author pairs appear only once in the final set.) The fact that more than a third have the "correct" indicative word in their titles is strong evidence that the search is on target. However, many potentially relevant items are retrieved that do not have *attribution* in their titles and that would not be retrieved in a search using that term. Furthermore, a search simply on *attribution* as a title word (or stem) would have drawn in more than a thousand items not related to attribution theory.

There seems considerable reason to believe that the cocited author search produced many good retrievals that could not have been obtained in any other way—certainly not by confining the search to PsycInfo. The PsycInfo indexers do not routinely capture all the articles relevant to or on a topic, as indicated by citations, under the appropriate descriptor (e.g., Attribution). Even if they did, Social Scisearch has its own distinctive journal coverage and is the most current of all databases in the social sciences. That of course makes it a useful complement to any other social science database when high recall is important.

The retrieval on Attribution from Social Scisearch alone amounts to a large literature. Very few would want to scan the entire list. Unless a comprehensive bibliography were the object (as might be the case), it would be simple enough to amend the search strategy in the usual ways—limiting output by language, document type, recency of publication, etc. It is also possible while online to experiment with different author combinations. Examples: Janis, an author less directly related to attribution theory than the others, adds least to the search and could be deleted. Or authors could be entered as ANDed triplets rather than pairs. This constrains output to papers citing all three, which may improve precision.

FREE PAIRING OF RELATED AUTHORS

Figure 2 lists authors used in an unusual retrieval in Social Scisearch, carried out in November 1982. The object was to obtain addresses of writers who had contributed theoretical or empirical studies to the literature of Science and Technology Indicators (S&TI). The addresses were wanted by the Center for Coordination of Research on Social Indicators in Washington, D.C., in part to prepare for a conference on measurement of science and technology. Writers on the many aspects of this topic were potential invitees; hence the need for addresses. The multidisciplinary S&TI literature sprawls over many journals and cannot readily be retrieved with subject descriptors. Robert Pearson, of the Center staff, had the idea of using cocited author retrieval, as demonstrated in an article by White and Griffith,[8] to obtain the wanted list. He supplied the two columns of authors on whom the search was to be based.

The choices in figure 2 were not entirely a matter of free judgment. The column at left comprises authors who were mentioned three or more times in the text of a U.S. government publication, *Science Indicators 1980*. On the right are authors mentioned at least five times in the index to *Toward a Metric of Science*, a collection of essays by leading interpreters of the science indicators movement. Both columns contain some well-known names. Those in the right-hand column are generally sociologists, historians, philosophers, and information scientists, while those on the left include science indicators specialists per se and a number of economists concerned with quantifiable effects of technological innovation. To both lists Pearson added names of other authors not drawn from the two "source" publications but with oeuvres relevant, in his view, to their contents. (These are labeled "judgmental additions.")

Papers to be retrieved were those that cited at least one author from each of the two columns. In other words, they could

From *Science Indicators 1980*	From *Toward a Metric of Science*
Frank J. Atelsek	Bernard Barber
Mark Carpenter	Joseph Ben-David
Edward E. Denison	Albert D. Biderman
Bela Gold	Harvey Brooks
Irene Gomberg	Jonathan R. Cole
Zvi Griliches	Stephen Cole
Christopher T. Hill	Otis Dudley Duncan
Edwin Mansfield	Yaron Ezrahi
Jon D. Miller	Eugene Garfield
Mary Ellen Mogee	Belver C. Griffith
Francis Narin	Warren O. Hagstrom
Keith Pavitt	Gerald Holton
Robert Pearson	Manfred Kochen
Kenneth Prewitt	William Kruskal
Anthony Romeo	Thomas S. Kuhn
Nathan Rosenberg	Imre Lakatos
Francis W. Rushing	Robert K. Merton
Frederic M. Scherer	William F. Ogburn
Luc Soete	Michael Polanyi
Lowell Steele	Karl Popper
Koray Tanfer	Derek de Solla Price
Nestor Terleckyj	Henry G. Small
James Utterback	Stephen Toulmin
Raymond Vernon	John Ziman
Samuel Wagner	Harriet Zuckerman
Judgmental Additions	
Frank M. Andrews	Diana Crane
Daryl Chubin	Christopher Freeman
Robert Evenson	Richard Freeman
Jerry Gaston	David Landes
Loren R. Graham	Alan Mazur
Todd LaPorte	Robert S. Morison
J. Scott Long	Nicholas Rescher
Michael J. Moravcsik	Peter Weingart
Nicholas C. Mullins	
Harold Orlans	
Donald C. Pelz	
Dorothy Zinberg	

Fig. 2. Science and Technology Indicators: Cocited Author List.

cite any pair or any combination of pairs of the seventy authors as long as one member was from the thirty-seven authors at left and the other was from the thirty-three at right. (Note that this differs from forming all possible pairs of the seventy. Pairs within columns, such as Atelsek and Carpenter, were not formed.) There was no attempt to create each pair separately, as in the attribution theory example. Given DIALOG's 99 set limit, this cannot be done and would be horrendously expensive and boring—37(33) = 1,221 pairs!—even if the limit were lifted. What was done, however, had the same effect as forming all 1,221 pairs separately.

The search strategy was in fact quite simple. Sets were formed on all seventy authors as Cited References. All those in the left column were then ORed together, exactly as if they were synonymous subject headings. The same was done with all names in

the right column. DIALOG's shortcut ORing, which allows the union of up to twenty-nine sets, was used. (For example, "COMBINE 1-29/OR" unites the sets of papers citing the first twenty-nine authors in the left column.) This tactic was continued until papers citing authors in the left column were one set and papers citing authors in the right column were another.

Predictably, since the lists include some very heavily cited authors—Merton, Kuhn, Duncan, Griliches, etc.—the final two sets were very large: 9,917 and 19,696 items respectively. (Almost surely they were inflated somewhat with homonymic retrievals, from names like CR = LONG J? or CR = ANDREWS F?) The final step was to AND the two sets together, which yielded a retrieval of 1,526 papers (Presumably these included few or no papers citing the "wrong" J. Longs, F. Andrewses, etc.). All retrieved citations were printed offline in Format 5 (Full Record), which in Social Scisearch provides not only the addresses of authors, as desired, but also the full cited references of each paper—useful as a check in judging relevance.

In her article on cocitation searching[9] Knapp mentions a Chinese menu strategy: "a 'hit' must contain, at least, one from column A and one from column B." The version of this strategy presented here may seem more like renting the whole restaurant for a fifteen-course meal. Be that as it may, the constraining effect of one column on the other worked much like that of one author on another, as in the example of Arendt and Nolte above.

I made the initial judgment on hits in the 1,526-item retrieval. Slightly more than half (791) appeared relevant to "Science and Technology Indicators," broadly construed. Considering the looseness of the topic and the lack of usable subject indexing, a precision ratio of about 50 percent seems acceptable. More rigorous judgment of "hits" would lower precision, but there is little doubt that many S&TI writers were found who would not have been discovered in a more conventional retrieval.

Interestingly, the search retrieved papers from not one but several literatures. I categorized these, based on chapters in a comprehensive review,[10] and assigned papers to them as in table 1. Assignments are based on titles, discipline or affiliation of authors, and writings cited. Papers on science made up 46 percent of the retrieval; on technology, 34 percent; and bibliometrics papers, 20 percent. The clearest examples of S&TI papers are found in the top six categories; the historical and philosophical papers seem more peripheral. "Predictable" journals that turn up include *Scientometrics*, *R & D Management*, *Technological Forecasting and Social Change*, *Futures*, *Adminsitration and Society*, *Technology and Society*, *Journal of Research Communication Studies*, *Journal of the American Society for Information Science*, and *Research Management*. The scatter, however, is very great, and it is in assembling papers and authors from "unpredictable" journals that this search shows its particular value.

Table 1. Yield on Free Pairing of Cocited Authors: Science and Technology Indicators.

Topic	Items Retrieved
Sociology of science	234
Bibliometrics	156
Economics of technology	98
Technology policy	97
Science policy	77
Sociology of technology	38
History of technology	28
History of science	25
Philosophy of science	21
Philosophy of technology	10
Economics of science	7
	791

Including supplementary retrievals on title terms (such as SCIEN? AND INDICATOR?,) and on known cited documents (such as *Toward a Metric of Science*), the total cost for online searching and a massive amount of offline printing came to about $370. (A discount was granted because my university subscribes to printed versions of the ISI citation indexes.) This sum is not a great deal to pay for a mailing list. If many authors' names are new to the list, that is valuable; and if they are not, one is confirmed in thinking that a certain "universe" is pretty well known, which is valuable also. In addition to serving practical ends, however, this cocited author retrieval made a good beginning toward a large-scale S&TI bibliography that is potentially publishable.

REFERENCES

1. Howard D. White. "Cocited Author Retrieval Online: An Experiment with the Social Indicators Literature." *Journal of the American Society for Information Science* 32:16–21 (Jan. 1981).
2. Janet Chapman and K. Subramanyam. "Cocitation Search Strategy." in *Proceedings of the National Online Meeting, 1981.* Medford, N.J.: Learned Information, 1981. p.97–102.
3. Howard D. White and Belver C. Griffith. "Author Cocitation: A Literature Measure of Intellectual Structure." *Journal of the American Society for Information Science* 32:163–71 (May 1981).
4. Howard D. White and Belver C. Griffith. "A Cocitation Map of Authors in Judgment and Decision Research." In *Concepts in Judgement and Decision Research: Definitions, Sources, Interrelationships, Comments,* B.F. Anderson and others, eds., New York: Praeger, 1981, p.261–71.
5. Howard D. White. "A Cocitation Map of the Social Indicators Movement." *Journal of the American Society for Information Science,* 34:307–12 (Sept. 1983).
6. White, "Cocited Author Retrieval Online," p.16.
7. *Concepts in Judgement and Decision Research: Definitions, Sources, Interrelationships, Comments,* B.F. Anderson and others, eds., New York: Praeger, 1981, p.14.
8. Howard D. White and Belver C. Griffith. "Authors as Markers of Intellectual Space: Cocitation in Studies of Science, Technology, and Society." *Journal of Documentation* 38:255–72 (Dec. 1982).
9. Sara D. Knapp. "Cocitation Searching: Some Useful Strategies." *Online* 8:43–48 (July 1984).
10. Ina Spiegel-Rosing and Derek de Solla Price, eds. *Science, Technology and Society: A Cross-Disciplinary Perspective.* London: Sage Publications, 1977. ■■

Library Applications of Electronic Imaging Technology

Clifford A. Lynch and Edwin B. Brownrigg

INTRODUCTION

The various technologies of electronic imaging—specifically, optical disk, document digitizers and Group IV Fax machines, laser printers, and bit-mapped displays—have now reached a critical stage of refinement and cost performance. So far, these technologies have seen minimal use in the library environment (except for a few pilot projects, such as the one at the Library of Congress.[1] We argue that this situation is about to change, and in fact these technologies not only offer new opportunities to extend the current developments in computer-based information delivery, but also can provide solutions to a number of very costly and long-standing management problems facing major research libraries. Additionally, we believe that the scale of these problems (and the financial stakes in solving them) is not well known.

This paper deals with three specific areas and attempts to give an overview of the magnitude of each problem, the applicable technologies, and the further technological developments that are necessary. In all cases, we draw heavily on data about the University of California's nine-campus library system, which, considered as a whole, represents one of the world's greatest research collections. The concerns we describe—storage requirements, preservation of library collections, and remote access to library resources (both in the sense of delivery to the patron and as resource sharing among cooperating libraries)—are, however, common to most major research libraries.

STORAGE SPACE REQUIREMENTS

The University of California needs about twelve new linear miles of shelf space annually to house its new acquisitions.[2] This implies the need for continuing major construction programs of enormous cost. Even if the university could afford an ongoing construction program of this magnitude, there are considerable difficulties in simply locating real estate on the various campuses to house new or vastly enlarged libraries. The current solution is to build regional storage facilities (essentially warehouses) for older or less-used material. One such facility is operational in northern California, and a second is under construction in southern California on the UCLA campus. The northern facility, located in Richmond, contains some fifty miles of dense shelving and is extendable to offer much more. The normal mode of use is for a patron at a campus to identify and recall material to his or her campus, although it is possible for patrons to visit the facilities and browse materials stored there.

Two problems are implicit in the current state of affairs. First, maintaining the regional storage facilities, while less costly than building new libraries, is still expensive. In the future, continued expansion of the facilities will be required on a periodic

---

Clifford A. Lynch is assistant director for computing and telecommunications and Edwin B. Brownrigg is director of the Division of Library Automation, University of California.

This paper is an expanded and revised version of a talk given at Electronic Imaging 85 in Boston, Mass., October 7–10, 1985. An outline of the paper appears in the digest of papers that was published for the conference.

basis due to the inexorable requirements for ever more shelf space. Second, the lower-cost regional storage facilities have substantially reduced access to the portion of the collection stored there.

The long-range solution is to digitize this material (either as images or, as the technology grows sophisticated, a mixture of text and images) and to deliver the material over telecommunications links both to printers in the campus libraries or elsewhere on campus and to end-user personal computers or work stations with high resolution, bit-mapped displays. The delivery of materials via telecommunications links is an increasingly desirable and feasible method of improving service quite independently of the storage problem and will be discussed later.

CONVERSION STRATEGIES FOR LIBRARY COLLECTIONS

Given the size of the University of California's collections, substantial conversion of even a portion of this material to electronic form will not be completed quickly. It will be necessary to provide for a partially converted operating environment that can be expected to persist for more than a decade and to devise policies that will identify and prioritize material to be converted to digital images.

There are two approaches that can be taken to digitizing the material at the storage facilities. One method is the wholesale digitization to optical disk, with delivery from the digitized archives on demand. This solves the space problem (ultimately making the current warehouses unnecessary) and provides fast, convenient delivery of material. There are drawbacks to this mass conversion. It would be costly due to the large amount of material being scanned. It also would commit major resources to the conversion of material that is by definition rather lightly used, rather than emphasizing digital conversion of the most popular material (within the overall collections of the University of California, not just those housed in the storage facilities) to facilitate convenient delivery.[3]

A second approach addresses the delivery problem for warehoused materials. Although it does not significantly affect the space problem, this method is considerably cheaper. Material in the warehouse would be digitized on demand and delivered electronically. Over time, this approach may greatly reduce the amount of manual processing needed to deliver material from the warehouses. Journals occupy a great deal of the shelf space in libraries, and chronological runs of old journals are among the prime candidates for storage at the warehouse facilities. In many fields it is common to find a rather small set of key journal articles that are of continuing interest and often requested. Current policies for selected material to be stored do not allow this degree of precision, and the simple fact that many issues of a given journal are typically bound together makes the "quanta" of storage inconvenient for retaining small numbers of popular articles on the campuses. But by storing a digital copy of each article that is digitized on demand, a library of these key articles would be built up, and repeated requests for these articles could be satisfied directly from the optical disks without further human intervention.

There are complex copyright problems that have to be dealt with before either approach can become a reality, however, and the economic and legal considerations may ultimately prove to be more difficult than the technical problems.[4]

There are cases, such as massive archives of meteorological observations (to which copyright does not apply) in which optical character recognition (OCR) technology rather than straightforward digital image capture may be effective. The value of such data would be greatly increased, since it would be in machine-readable formats suitable for statistical analysis after OCR processing. For textual material OCR is favored generally over imaging because it directly captures the content of the material. This opens the prospect of computer-directed indexing. Character-form data also requires far less storage and transmission capacity and can be delivered on any low-cost terminal or printer. However, except for special cases where page layouts are simple and regular and fonts are well-controlled, really reliable and cost-effective OCR technology remains tantalizingly just beyond the state of the art. One

major difficulty is the lack of well-accepted methods of capturing the meaning implicit in multiple font styles (italic, boldface, etc.), type sizes, and the arrangement of text on a page as part of the OCR process.

PRESERVATION AND CONSERVATION

The material in a library's collection deteriorates over time. Some of this deterioration is just the wear and tear of normal use or occasional vandalism. More devastating, however, are the effects of almost two centuries of publication on acid paper. Environmental damage has also taken its toll. A 1983 University of California report indicates that roughly 16 percent of the UC collection (some three million volumes) are now so brittle that the volumes are difficult to handle; another three million are badly deteriorated.[5]

The two preservation methods currently used within the library community are mass deacidification, in plants where books are treated with various chemical processes to eliminate the acid in paper, and microfilming of selected books. Both of these approaches are expensive, and libraries are working cooperatively to eliminate duplicate processing (especially in microfilming) in an attempt to minimize costs. Aside from cost, microfilming makes material inaccessible. It generally does not circulate; and even when it is available for loan, few patrons have microfilm readers. Microfilm is cumbersome to browse and difficult to read. And most devices for producing paper copy from microfilm yield nearly illegible prints. Microfilm is almost universally hated by library patrons, who will avoid using it as much as possible. If copies of the microfilm are not stored at all major research libraries, there is also a delivery problem. Furthermore, microfilm itself will deteriorate over time, especially with frequent and sometimes abusive handling.

The deterioration of library collections is perceived as an enormous threat, and indeed it is. However, if digitization to optical storage is substituted for microfilming and perhaps for deacidification as well, it becomes a great opportunity to make collections more, rather than less, accessible. Once digitized, copies of material can be sent from one library to another across telecommunications networks. Space for storage is reduced, and the material is in a form that permits direct delivery to end users. Best of all, much of the older material is now in the public domain since copyright protection has expired, simplifying legal and administrative problems.

Interestingly, recent experiments at the National Archives[6] suggest that algorithmic enhancement of digital images actually can be an effective way to compensate for a wide range of damage—fading, water staining, etc.—that plagues older material.

One can go a step further and consider the possibility of imaging very rare and valuable books, manuscripts, and other material currently held by major research libraries. Libraries, much like museums, have carefully maintained and preserved their collections of such material. And with museums, they share the problem of collection accessibility. Because of its great values, rarity, and fragility, much of this material is currently accessible only to advanced scholars and cannot be used by students. Imagine the results of making copies of original source material—ranging from presidential diaries to the Domesday Book or the Gutenberg Bible—available to students as part of a basic undergraduate curriculum. In many cases, these national treasures may be best distributed by publication on videodisc or CD/ROM, however, rather than across telecommunications networks.

REMOTE ACCESS: THE ELECTRIC LIBRARY

Research libraries are developing very sophisticated computer-based online catalogs to replace their traditional card catalogs.[7] These can be accessed twenty-four hours a day from home and office terminals, so it is no longer necessary to go to the library to *identify* material.[8] At present, however, material cannot actually be *delivered* through telecommunications media. End-user demand for remote access is growing. Solving this problem again requires optical disks to house the material, document scanning to capture material that is not available in machine-readable form, and personal computers with bit-mapped displays and laser printers to pro-

vide output devices.[9] Significant amounts of local storage (perhaps using optical write-once media) are also needed by the end user to maintain personal "libraries" of material. The availability of such local storage, in conjunction with appropriate end-user software, may eliminate much of the disorganized paper that currently engulfs many researchers.

The electronic delivery of information also offers other important benefits for libraries. It creates greatly expanded opportunities for coordinated acquisition of materials by cooperating groups of libraries, since a single "copy" can be delivered quickly anywhere across a telecommunications network. This, of course, again raises issues of copyright and could totally alter the economics of journal publishing and purchasing.[10]

In the longer run, however, the benefits of storing and delivering electronic material rather than paper copies go far beyond quick delivery though telecommunications. Ultimately, conversion of library collections to electronic form is an essential step toward controlling the explosion of published literature that today overwhelms most library patrons. The first applications of digitization will employ simple image digitization techniques. As the technology advances, much material will be scanned through optical character recognition rather than imaging techniques. (It is interesting that the proposed Group IV FAX standards for Class 2 and 3 machines explicitly incorporate the ability to process mixed image and text data.) At the same time, more and more newly published books and journals will become available in machine-readable (character) form. The availability of large amounts of machine-readable material will permit libraries to enhance their online catalogs (which currently use databases consisting solely of descriptive bibliographic data) to also include actual information from books such as tables of contents, introductions, and indexes. These entries can be scanned online by the patron to more precisely identify material of interest. In addition, computer-based indexing, filing, and retrieval systems based on expert systems technology will be able to process such a database of text. This will permit more extensive and exhaustive indexing and classification than current labor-intensive, computer-assisted methods allow. One very interesting prospect is the ability to revise older cataloging and indexing in light of later developments in a field, thus producing databases that can be searched as a consistent whole, rather than as an agglomeration of many historical layers of less than consistent cataloging and indexing.[11]

If we look further into the future, wholesale digitization of materials (other than diagrams, pictures, etc.) will become less commonplace. Prospective material will be available in machine-readable (character) formats. Older items that have been digitized will be largely converted to character format via OCR techniques applied to existing digital images of the material (or direct OCR scan, when these items still exist in paper form).

Thus, we would argue that while electronic imaging technologies may be central to libraries over the next fifteen years, they will ultimately become intertwined with and largely overshadowed by optical character recognition technology, very likely augmented with computer-based automatic indexing and filing systems.

BARRIERS TO THE IMPLEMENTATION OF IMAGING SYSTEMS

Technology

Most of the basic technology exists to allow a start in implementing the applications we have discussed. What is missing is largely *systems integration:* the software, standards, and packaging to make the equipment operate and interoperate effectively for the proposed applications.

Laser printers with very high resolution are readily available at lower and lower costs. High-resolution displays, too, are progressing, but more slowly. However, protocols and data formats for transmitting images or mixed text and images to these devices are not well established. The Group IV FAX standard is one attempt at this, but falls short in several respects; the applications under consideration here go far beyond the goals of the FAX standards. Several relevant officially sanctioned or de facto standards-making efforts are under-

way: IBM's DIA/DCA, the Association of American Publishers' Electronic Manuscript Project, and work within ISO. Once standards are agreed upon, widely available software to implement them will be required.

While image scanning has made enormous progress recently, burdensome operational aspects remain. For example, most current imaging experiments require quality control checks for almost every page scanned. The integration of scanning and optical character recognition may ultimately permit more self-adjusting scanners, since the OCR component can control the scanner parameters and attempt rescanning.

Optical disk technology is well developed and promises increasing storage densities in the coming years. However, the hardware and software integration of optical disks with current large, state-of-the-art mainframes (such as large IBM systems) lags far behind the raw capabilities of optical storage proper.

OCR technology, the key to many of the most exciting long-term prospects, still requires considerable development. However, a great deal can be done prior to the maturity of OCR using pure imaging techniques. Additionally, there is a real need for pilot projects using OCR technology.

Cost Factors

Obviously, imaging technology must be affordable and cost-effective to succeed beyond the experimental stage. This means that not only must the costs of the new technology compare favorably with the existing solutions, but also that the technology must be priced to permit gradual introduction (rather than impractical sudden and total conversion). Given the relatively small quantities necessary, scanners, printers, and optical disks already meet these criteria.

The two greatest cost factors in electronic imaging are in scanning and in the delivery system.

Scanning today is labor intensive; the scanners require manual feeding and quality control checks. Although it is very difficult to apply technology to solve these particular problems, there will be a substantial payoff for significant advances. Ultimately perhaps, advances in robotics technology will yield some progress in this area.

Practical document delivery presupposes the existence of an infrastructure—very high speed networks and large numbers of high-resolution bitmapped displays—to support it. High-speed networks are starting to appear as local area networks and fiber optics become more commonplace, but large-screen, high-resolution bit-map displays remain expensive and relatively uncommon. Moreover, these displays are normally integrated with personal computers, creating the need for software to receive text and images from the network and drive the display. Rewiring university campuses to provide high-speed local area networks is a slow, enormously expensive undertaking but it is inevitable, since computing and telecommunications have become integrated with all areas of academic research. The most cost-sensitive component of a document delivery system is the high resolution display. While the costs of local area networks can be amortized across a very wide range of applications, the immediate importance of very high resolution displays (and the computing power and memory to support them) is more limited. Given the large numbers of displays that will be needed to make widespread end-user document delivery a reality, the rate at which such delivery becomes practical is closely tied to the costs of such displays. An institution the size of the University of California could easily absorb tens of thousands of such displays; in these quantities, each $100 reduction in the unit cost of such technology reduces the overall investment necessary to develop the infrastructure for document delivery by millions of dollars.

Other Issues

Finally, we must keep in mind the nontechnological barries to the replacement of warehouses of paper with electronic images. The economic structure of the publishing industry is not yet prepared to cope with this change. Congress and the courts are struggling with the application of copyright in an electronic environment. And there is great psychological resistance to the elimination of paper and (sometimes well-

founded) distrust in the reliability of electronic or optical storage as an archival medium by both librarians and library patrons.

The technology of paper has been refined for centuries. As a means of placing most kinds of information in the user's hands, it is still very effective. While there has been speculation about the paperless world of the future, we believe that the major changes will be in the amount of paper that is stored and physically shipped. Low-cost, high quality printers are an essential part of imaging technology, and we expect them to be heavily used. In fact, the amount of printing may actually remain constant or increase with the introduction of imaging systems, but it will be single-copy printing on demand.

CONCLUSIONS

The first two problems we have discussed—the unending demand for storage space and the deterioration of library collections—are primarily problems of library management. They affect the library user only insofar as very real budgetary constraints have forced libraries to sacrifice access to material in order to solve these problems. In the end, the result is a slow, incremental deterioration of service.

The third class of challenges revolves around various aspects of greatly improved ability to identify and obtain materials, and these form the heart of library service. The consequences of meeting these challenges may have far more impact than simply stemming a gradual decline in service. What is at stake is fast, widespread, and accurate access to information.

ACKNOWLEDGMENTS

We would like to thank Nancy Gusack for her editorial assistance.

REFERENCES

1. D. F. Willis. "Image Processing Using Optical Disk Technology at the Library of Congress," in *Optical Disks Systems and Applications*, Edward V. LaBudde, ed., *Proceedings SPIE 421* (1983), p.92–95.
2. University of California, Office of the Executive Director of Universitywide Library Planning. *The University of California Libraries: A Plan for Development, 1977–1988* (Berkeley, Calif.: Univ. of California, 1977).
3. Clifford A. Lynch and Edwin B. Brownrigg. "Conservation, Preservation and Digitization," *Proceedings of the ACRL Fourth National Conference, April 9-12, 1986, Baltimore Maryland* (in press).
4. Edwin B. Brownrigg and Clifford A. Lynch. "Electrons, Electronic Publishing, and Electronic Display," *Information Technology and Libraries* 4:201–7 (Sept. 1985).
5. University of California, Office of the Executive Director of Universitywide Library Planning. *Conservation of the Collections: A Supplement to the University of California Libraries: A Plan for Development, 1977–88* (Berkeley, Calif.: Univ. of California, 1983).
6. William L. Hooton. "The Optical Digital Image Storage System (ODISS) at the National Archives and Records Administration," in *Proceedings of the Videodisc, Optical Disk, and CD-ROM Conference and Exposition, Philadelphia, Pennsylvania, December 9-12, 1985* (Westport, Conn.: Meckler, 1985), p.90–94.
7. Charles R. Hildreth. *Online Public Access Catalogs: The User Interface* (Dublin, Ohio: OCLC, 1982).
8. Edwin B. Brownrigg and Clifford A. Lynch. "Online Catalogs: Through a Glass Darkly." *Information Technology and Libraries* 2:104–15 (Mar. 1983).
9. Clifford A. Lynch and Edwin B. Brownrigg. "Document Delivery and Packet Facsimile," *Proceedings, 48th ASIS Annual Meeting, las Vegas, NV, October 1985* (White Plains, N.Y.: Knowledge Industry, 1985), p.11–14.
10. Edwin B. Brownrigg, Clifford A. Lynch, and Mary E. Engle. "Technical Services in the Age of Electronic Publishing," *Library Resources and Technical Services* 28:59–67 (Jan./Mar. 1984).
11. Clifford A. Lynch. "Catalog Practices and the Online Catalog," in *Proceedings, 48th ASIS Annual Meeting, Las Vegas, NV, October 1985* (White Plains, N.Y.: Knowledge Industry, 1985), p.111–15. ■■

Library Automation in the People's Republic of China

John H. Maier

GENERAL INTRODUCTION

This article is the result of visiting more than sixteen of the top university libraries, various Chinese Academy of Sciences libraries, and the National Peking Library—the equivalent of the Library of Congress—during a year of work, research, and lecturing in the People's Republic of China during the 1984–85 academic year. Discussions were held with librarians, administrators, and computer science personnel, with a specific interest in Chinese progress in and plans for automation of libraries and with a view towards China's entrance into the global age of information technology.

Why should we in the United States be interested in library automation in the People's Republic of China? Library science essentially concerns itself with the acquisition, ordering, storage, and availability of knowledge. Only recently has this involved computerization as a tool. As in other disciplines, however, computerization has changed the work of librarians as perhaps no other tool used by the discipline. As one critical effect, library automation has opened up the vistas of networking, so that the whole is becoming greater than the sum of the parts. There will come a time when it will be possible through online capabilities to access all of the important libraries of the world, and libraries in the People's Republic of China will be among them.

In the future, though it may be a decade, decades, or more, it will be possible to access machine-stored information from libraries anywhere in the world and in any language. The technical, organizational, and political barriers to this possibility are enormous, but in part the problem is that we are viewing the real possibility from this side of the paradigm, with much of the challenge still in front of us. It will happen faster in some parts of the world than others, and, in fact, the United States could be considered a world leader; but the trend in countries everywhere is slowly emerging, certainly in Europe, now in Asia, and in time elsewhere.

A plethora of world languages makes this concept interesting. However, right now, the European Economic Community is working hard to implement automated translation systems for its primary languages. For Chinese there are current research projects in the United States, China, Hong Kong, England, and elsewhere that are making substantial progress. Someday it will be possible to ask a system, networked to other systems, for all materials on *Boolean algebra*, for example, and the answer will include a list that cites holdings in all major world libraries, in all major languages, and perhaps even in some obscure ones.

The reason, then, for our interest in library automation in the People's Republic of China is that someday it may affect the range, the potential, and the quality of our library services in the United States.

LIBRARY AUTOMATION IN THE PEOPLE'S REPUBLIC OF CHINA

In 1978 the vice-premier for science and

John Maier lectured as a professor of computer science in China for one year. His special interests are computer networking, information retrieval, and machine language translation, with an emphasis on the economic and social impact of information technology.

technology, with the support of China's highest leaders (Deng Xiaoping attended the conference), established "automation of key research centers" and "national networking" as top priority goals in China's modernization and science and technology plan. By 1985 little implementation progress had been made. To date not a single library in China is fully automated,[1] and there are eight hundred Chinese universities and colleges that might be considered candidates for automation. To see exactly what is happening we need to look closer.

In 1978 China faced considerable hurdles to automation and national networking of research centers. The difficulties essentially involved five areas: hardware inventory; software applications development; Chinese character (ideograph) automation, including input/output and internal processing standards; library classification system(s), and networking topology. By 1985, however, significant progress had been made in some of these areas, and in one—computer processing of Chinese ideographs—important technological breakthroughs had been made. Now the necessary ingredients for China to automate its information resources, network them, and also link to other global systems,[2] are, or will soon be, available. Each of these five ingredients, which are actually generic, will be discussed in detail.

HARDWARE INVENTORY

In 1978 China was still emerging from the ravages of the Cultural Revolution. Universities had only reopened a few years earlier after a decade of loss. Domestic computer production capabilities and designs, though not inconsequential, were dated and were mainly second-generation, perhaps early third-generation machines, largely devoted to scientific applications. Nonscientific (nonengineering and nonmathematical) computer applications were almost unknown in China.

By 1979, however, and increasingly so thereafter, China began to import hardware from the West and began to consider more general computer applications. Western export control laws began to reflect the status of China as a friend and significant trading partner with the West, not as an adversary. China's universities started building an inventory of hardware, at first microcomputers. With a World Bank loan of $250 million in the mid-eighties for university automation projects, a large number of universities acquired IBM, Honeywell, DEC, Data General, and other good-sized computers. Almost universally now, Western, usually American, and some Japanese computer products prevail on campuses. For example, one manufacturer, Honeywell, has computers in almost every top Chinese university.[3] All of the computers can support approximately sixty terminals. Though Chinese campuses do not yet have the cornucopia of computing power found on American campuses, installations now in place are sufficient for immediate needs.

SOFTWARE APPLICATIONS DEVELOPMENT

China still lacks broad experience in management of nonscientific applications design, specification, development, implementation, and maintenance. Software engineering is itself an art as well as a science, and its management, large-scale coordination, and scheduling require thorough knowledge of the software development, integration, and test process, as well as knowledge of how to successfully transfer new automation procedures into old organizational structures. China has a dearth of well-trained programmers, analysts, systems consultants, and other computer professionals that are needed so that projects and priorities compete for the talents that are available.

Libraries in the West have experienced the pain of converting to automated systems, with the concomitant reorganization of job functions and necessarily thorough evaluation of work flow that automation requires. To China's benefit there is now a wealth of literature and experience on which to draw, and well-established library automation techniques and standards will save China's library automation teams time and effort. In some cases it may even be possible to migrate developed systems already in use elsewhere to China's computers. However, at this time software engineering and management are a signifi-

cant concern for Chinese planners.

CHINESE IDEOGRAPH AUTOMATION

In 1979 automation of Chinese ideographs seemed to be the single largest, and perhaps insurmountable, barrier to automation of China's libraries. In the West, using the standard QUERTY typewriter keyboard, we don't even give much thought to our relatively small character set. For the familiar character set of upper- and lowercase *A* thru Z, numerics, and special characters, most computers now use the standard ASCII (American Standard Code for Information Interchange) binary representation (8 bits), which enables compatibility between systems—an *A* on one system is represented by the same binary string as an *A* on another system. China's ideograph set, however, comprises more than forty thousand unique characters, with perhaps three to four thousand in most common use. For even a start, the question was how to design an efficient input keyboard.

Nevertheless, by 1985 both input and output of Chinese ideographs, as well as internal processing, had essentially been solved with different systems, actually numerous, available on the market in China and in the West. The question became only one of choice. China's State Council, in 1979, established a binary string (16 bits), standard representation for Chinese ideographs.[4] The result is that it is now possible to combine Chinese and alphanumerics internally and to use their binary representation in multisystem compatibility, which then facilitates interlibrary and international networking.

CHINA'S LIBRARY CLASSIFICATION SYSTEM

China has its own library classification system, and in fact at least two are in widespread use—one promulgated by the Chinese Academy of Sciences (CAS) and one by the Peking National Library, China's equivalent of the Library of Congress.[5] The latter is most widely used by university libraries. Thus, China does not presently use the International Standard Book Number (ISBN), the International Standard Serial number (ISSN), the Dewey Decimal System (DDS), or the Library of Congress Call Number (LCCN), all of which are widely used elsewhere for classification of publications and of subject information. However, it is slowly making movements in the direction of some of these systems, the ISSN, for example.

China's systems are dated, inadequate, and in some areas inappropriate. For example, I found that categorizations within computer science did not reflect state-of-the-art activities worldwide. Reform of China's library classification system(s) will be necessary in the long run, but automation can nevertheless proceed now. Future changes can be handled through software conversions or other techniques. In fact, system software design should always maintain the dynamic nature of information and the ability to absorb changes and conversions to other classification paradigms.

NETWORK TOPOLOGY

In the United States we now take for granted pervasive, high-quality communications services and support, but China is a developing country where 50 percent of all phone calls are not completed and circuits are not up to standards required for computer data transfer. Right now, telecommunications has the highest priority in China's modernization plans. When China does implement a network of research centers, it will be a first.

Networking technology requires both software and hardware, at the host CPU, and within a reliable communications infrastructure. China has recently acquired a geosynchronous communications satellite launch capability and is discussing cooperation in this technology with the United States. China need not follow the earlier model of the West by laboriously building terrestrial communications and can, like other developing countries, enter the global village of information technology with the launch of a single advanced satellite.[6]

China can now realistically begin to plan implementation of a national data communications network to link automated libraries. International connections will follow. A project of this sort, even in the United States, would be a challenge for any manager. China must first identify a test-

bed group of geographically dispersed, and important, libraries for prototype implementation, and assign high priority to concurrent scheduling and compatibility of its library automation activities while ensuring that necessary support communications hardware is procured, installed, and tested. A "network control" will need to be established to ensure effective coordination—any number of models are available in the West. Cooperation between institutions will have to be well managed.[7] China can use the International Standards Organization standard X.25 packet, switching applications layer ANSI Z39.2, ISO 2709 standard for a message format between host applications and thus ensure compatibility between disparate IBM, Data General, DEC, Honeywell, Fujitsu, and other computers now in China. Finally, ALOHA broadcast would be an appropriate network topology for China.[8]

SUMMARY: WILL YOUR LOCAL LIBRARY BE ONLINE WITH PEKING?

Yes, the day will come when your local library will be online with Peking, and Tokyo, and Paris, and London, just as today it is possible to communicate with those locations by letter. The trends of electronic information technology are clear,[9] and work in China has begun.[10] From a technical viewpoint the most important consideration is compatibility of standards. Without agreeing upon a method for communicating, it is not possible to communicate.

> Without standards, a great deal of effort would be wasted in repeatedly designing and developing data structures, and so on, effort that could be more (productively) applied to solving basic human problems.[11]

China generally participates in the programs and activities of the International Standards Organization. For example, it can use the Machine-Readable Cataloging (MARC)[12] format, which was developed by the Library of Congress to facilitate compatibility for electronic exchange of information between libraries, is in widespread use, and complies with the International Standard of Bibliographic Information Interchange (ISBII) (also American National Standard for Bibliographic Information Interchange, ANSBII, on Magnetic Tape, Z39.2-1971, revised ANSI 1978, now ISO 2709). In its important libraries China's library scientists need to go beyond their heretofore circumscribed concept of library automation and think towards full systems services, to include

A. Technical Services Subsystem Modules
   1. Selection and ordering
   2. Cataloging
   3. Preparation
   4. Collection control
   5. Serials
B. Reader Services Subsystem Modules
   1. Circulation
   2. Reference
   3. Interlibrary loan
   4. Information retrieval
C. Library Management Subsystem Modules
   1. Accounting
   2. Personnel
   3. Procedures control
   4. Management information
   5. Inventory control

When that is done, China's libraries can begin to network. When that occurs, I can think right now of several important centers in the United States that will be interested in some electronic handshaking.[13] And once those centers network to China, then—via networks well established in the United States—our local library will be online with Peking.[14]

Finally, for those who do not think it can or will happen, I would like to share with you some instructive Tang Dynasty poetry.[15]

> The woods are black and a wind assails the grasses
> Yet the general tries night archery
> And next morning he finds his white plumed arrow
> Pointed deep in the hard rock

ACKNOWLEDGMENTS

The author would like to thank Qi Wang, director of the Orientalia Division, U.S. Library of Congress, for sharing this vision, and Thomas Cheatham, Harvard University, who led the first computer science delegation to the People's Republic of China many years ago under the auspices of the American Association for the Advancement of Science. Finally, I wish to thank many Chinese colleagues and hosts.

REFERENCES AND NOTES

1. One university—Qinghua University, the MIT of China—in Beijing has had for several years a limited computerized library inquiry system that library members can use themselves, but the system only searches local English holdings and is not networked.
2. Growing systems include the Ohio College Library Center (OCLC); the earliest and perhaps largest, servicing more than seventeen hundred member libraries and three thousand terminals in almost fifty states and Canada; the Research Libraries Information Network (RLIN); the Washington Libraries Network (WLN); the University of Toronto Library Automation System (UTLAS), the British Library System (BLAISE), and others. Stanford/Hoover, has, perhaps, one of the first systems for cataloging Chinese.
3. Some of the universities with Honeywell installations include Beijing, Qinghua, Jilin, Fudan (where President Reagan visited and spoke), Shanghai Jiaotong, Zhejiang, Nanjing, Huazhong, Xian Jiaotong, and others.
4. "China's National Standard" (GB 2312-80), Code of Chinese Character Graphics Set for Information Interchange (Primary Set: 6,763 Characters), Technical Standard Press PRC, Guoji Shudian (China Publications Center), P.O. Box 399, Beijing. For input a variety of techniques are available, some using special enlarged keyboards to compose characters from their components and some using coding techniques correlating from each character to a standard ASCII keyboard. The best techniques use the standard keyboard, since this makes available the full range of hardware on the market. Pinyin, the romanization system for phonetic Chinese, is a favored technique with the full backing of the government. For an interesting related article, see Joseph D. Becker, "Multilingual Word Processing," *Scientific American* (July 1984).
5. At the most general level the Peking National Library classification system (using letter delineations) is as follows. (Source: *China's Library Classification System, Zhongguo Tushuguan Fenleifa*, Peking Library, 1983).
   - A Marxism, Leninism, Thoughts of Mao
   - B Philosophy
   - C Sociology
   - D Politics
   - E Military Affairs
   - F Economics
   - G Culture, Science, and Education
   - H Language
   - I Literature
   - J Art
   - K History and Geography
   - N Natural Sciences
   - O Physical-Mathematical Sciences and Chemistry
   - P Astronomy and Geosciences
   - Q Biology
   - R Medicine and Hygiene
   - S Forestry and Agriculture
   - R Industry and Technology
   - U Communications and Transportation
   - V Aviation and Spaceflight
   - X Environmental Science

   Note: There is a fairly healthy amount of exchange, right now, between the United States and the People's Republic of China in library science. For example, at the Peking National Library, where an automation project is beginning, I met staff members who had received recent training at the Library of Congress in Washington, D.C. Also, a Fulbright recipient in China during 1984–85 was an American library scientist.
6. Global/technological village–thinking has reached China's Academy of Sciences. For example, see *The Abrupt Rise of the New Technology Revolution* (Xin Jishu Geming De Jueqi) (Bejing: CAS Pub. 1984).
7. A reasonable phasing for a national network might include

   Phase One: Beijing (Honeywell) and Qinghua (Honeywell) universities as a close local net with important libraries,
   Beijing Institute for Aeronautics and Astronautics (IBM) for its machine mix and important library,
   Fudan (Honeywell) and Shanghai Jiaotong (Honeywell) universities as a close local net with important libraries,
   Nanjing University (Honeywell) with important library,
   Sun yat-sen University (VAX) with important library;
   this would constitute a prototype national network.

   Phase Two: Other universities with Honeywell installations (Jilin, Zhejiang, Huazhong, Tianjin, Xian Jiaotong, Dalian, and University of Science and Technology);
   this would constitute a full national network for China.

Phase Three: Selected Chinese Academy of Sciences institutes, Beijing National Library, and other universities.

This would expand and mature the network. Also, in phases two or three international linkages could be established.

8. ALOHA topology, developed at the University of Hawaii, is an elegant broadcast communications satellite methodology for village level communications.
9. For example, to see the rapidity of the spread of this viewpoint note the translation into Chinese of John Naisbitt, *Megatrends: Ten New Directions Transforming Our Lives* (Da Qu Shi) (China Social Science Pub., Beijing, 1984).
10. For example, note: "Tibetan Classics Stored in Computer," in *Beijing Review*, Sept. 24, 1984.
11. James E. Rush, "Library Automation Systems and Networks," in Marshall C. Yovits, ed., *Advances in Computers*, V.21 (New York: Academic, 1982).
12. In China today there are several universities with library science departments. Every large municipality, and all good-sized cities have public libraries. In addition to national scholarly library science journals, almost every province has a library association and publishes a library science journal. Throughout this entire community of library professionals I noted widespread discussion of, and many journal articles about, adoption and utilization of MARC format.
13. For an immediate list of institutions that would logically be interested in networking with China's libraries we could consider Library of Congress, Harvard Yenqing library, the Stanford/Hoover Institution library, Berkeley University library, University of Michigan Center for Chinese Studies library, and the University of Hawaii/East-West Center library.
14. As an example of what is possible, Guangzhou (Canton) is right now online with Lockheed's Dialog.
15. Li Bai is known generally and loved by all Sinophiles as China's preeminent poet during the Tang dynasty (AD 618–907) cultural effloressence. The parallel of the arrow striking the rock—i.e., the target—is the high priority information technology need of China today. ■■

Retrospective Conversion: A Question of Time, Standards, and Purpose

Phyllis A. Valentine and David R. McDonald

INTRODUCTION

Retrospective conversion of catalog records into a machine-readable form is a topic that has attracted attention for more than ten years. With the ever-increasing move toward online catalogs, interest in conversion is intensifying.

While the various methods of retrospective conversion have been identified and described, the costs of such activity are not well known. Estimates for converting a record run from $.75 to $5.25 depending upon (1) the method of conversion and (2) the costs included in the estimate. This wide variation in costs is not insignificant, especially for libraries with large collections. In many cases retrospective collection of an entire collection can exceed $1 million. In such a situation a difference as small as $.05 in the per-record cost of conversion can have a substantial effect on the total cost of a project. If, for example, there are one million records to convert, a reduction of $.01 in the per-record cost translates into a reduction of $10,000 for the project.

This paper examines the factors that determine the cost of retrospective conversion, reports the results of a cost study at the University of Michigan Library, and introduces for discussion an alternative strategy.

FACTORS DETERMINING RETROSPECTIVE CONVERSION COSTS

Peters and Butler have described in detail the factors that shape the cost of retrospective conversion.[1] The factors include (1) the definition of conversion, (2) standards of acceptance, (3) method of conversion, (4) hit rate, and (5) standards for the creation of machine-readable records for nonhits. In determining the definition of conversion, libraries must decide whether they want to simply convert the existing text on a card into a machine-readable form or upgrade records to existing standards. The latter approach requires greater effort than the former. In the same vein, the question of what constitutes a match from a resource file also affects the amount of effort required. As Peters and Butler note, "An exact match can range from matching the shelflist card in all bibliographic details to matching the author and title only."[2] The costs associated with in-house conversion will differ from those charged by a commercial vendor or bibliographic utility.

While several methods of retrospective conversion are available, they all rely upon the same strategy—the matching of cards to existing machine-readable records in a "resource file" and the copying of matching records to the library's local system. The utility of this strategy is directly dependent upon the degree to which a resource file mirrors a library's collection. If 90 percent of a library's collection is represented in the resource file, only a small percentage of records will have to be keyed; if only ten percent is represented, the amount and extent of keying becomes more important. If a library is intent on converting its entire

Phyllis A. Valentine and David R. McDonald are in the Systems Office, University Libraries, University of Michigan–Ann Arbor.

collection, it is likely that some number of records will have to be keyed; costs associated with this task will depend upon the extent of keying and the standards used to create those records.

Another reason published costs for retrospective conversion vary so widely is because not all estimates include the same factors. Some estimates include direct costs only and others do not include the cost of post- or preediting, for example. Level of staff used to perform the same tasks also varies from institution to institution.

The University of Michigan is in the midst of planning for an online catalog and, as part of the preparation, has been examining the cost of converting the entire collection. The library holds more than two million cataloged titles. The following records in machine-readable format are available:

465,000 monographs
95,000 serials
3,000 scores

as well as smaller numbers of other machine-readable formats such as maps and microforms. This represents just 28 percent of the total collection and leaves approximately 1.5 million records to convert—a substantial and costly task. In order to obtain a better understanding of the costs, the library undertook a time and cost study of the current conversion efforts at Michigan.

COST STUDY

Methodology

The study was undertaken to gather data on the time and cost for each step involved in the current complete conversion of monograph cataloging records into machine-readable form using RLIN. The study was also intended to determine the average standard time and cost for an entire record to be converted. Conversion was done through RLIN and to full MARC and RLG standards.

The conversion process was analyzed in order to identify steps for measurement. The type of step, the method of measurement, and the staff performing the step are listed below. A *fixed* type means that the step is performed for every record; a *variable* type means that the performance of the step depends on the particular record.

STEP 1: Categorization of the shelflist cards prior to photocopying.
Type: Fixed
Method: Direct timing, in 15-minute observations; LC classification was noted.
Staff: ReCon student hourly

STEP 2: Photocopying of the cards onto work forms
Type: Fixed
Method: Diary, by one drawer's contents
Staff: Photoduplication student hourly

STEP 3: Sorting of the work forms into categories for pre-RLIN work
Type: Fixed
Method: Direct timing, in 15-minute observations; LC classification was noted.
Staff: ReCon clerical

STEP 4: Searching the Public Catalog
Type: Variable
Method: Diary, in 30-minute periods; LC classification was noted.
Staff: ReCon student hourly

STEP 5: Searching the Authority File
Type: Variable
Method: Direct timing, in 15-minute observations; LC classification was noted.
Staff: ReCon clerical

STEP 6: Tagging the record for inputting
Type: Variable
Method: Direct timing, in 15-minute observations; LC classification was noted.
Staff: ReCon clerical

STEP 7: Resolving problems prior to RLIN work
Type: Variable
Method: Diary, for 30-minute time periods; LC classification was noted.
Staff: ReCon clerical

STEP 8: RLIN work, divided into searching, fixed fields, variable fields, holdings, problem solving, writing RLIN ID on workform
Type: Fixed, but each individual step except Holdings & Writing of RLIN ID is variable.

Method: Direct timing, in 30-minute observations. Record type & LC classification recorded. Response time was not included except within searching.
Staff: ReCon clerical

STEP 9: Writing of RLIN ID on shelflist cards
Type: Fixed
Method: Direct timing, in 15-minute observations
Staff: ReCon clerical

STEP 10: Total population of records and combinations of work done prior to RLIN work
Type: Step was added in order to gather data for study.
Method: Diaries (codes written on completed work forms). LC classification noted.
Staff: ReCon clerical.

Except for steps 2 and 10, data was gathered at random times from April 19 through April 25, 1985. Step 2 was determined by records compiled in the photoduplication unit, and step 10 was run continuously until a pattern was seen. Sample sizes were determined from average monthly production figures or through sampling until patterns emerged. Time measurements included travel, sorting, and alphabetizing time if present in the time period.

Results

The unit of measure for RLIN was seconds (converted to minutes in the results); all other steps were measured in minutes. All observed times have been converted to standard time; all staff costs have been calculated to include benefits, sick time, vacations and holidays, etc. Library of Congress cataloging classification was considered and is presented separately in the various tables when significant differences were noted.

The time and cost for each step is presented in table 1; the level of staff is included because of its influence on the cost. Those steps that are variable according to the record being handled are marked with an asterisk. RLIN work includes all steps from searching to writing down the ID number after inputting. As expected, those steps that involve the greatest amount of judgment and training (*searching the authority file, tagging the record, resolving problems,* and *RLIN work*) require higher level staff, take the most time, and cost the most.

Work on RLIN consists of six steps: searching, filling out fixed and variable fields, creating holdings, solving any problems (usually relating to tagging), and writing down the assigned RLIN identification number. The time required for three of these steps varied according to the *type* of RLIN record being entered.[3] Variation by cataloging classification was seen, but did not appear to be significant. Searching took the most time when the final result was negative (*Create*) or when the match was for an edition (*Create* *). The fixed and variable fields demonstrated the same pattern of more required time for *Create* and *Create* * records. Variations in the other three steps appear to be more a result of individual cataloging record variations than of either classification or RLIN record type. Table 2 outlines the variations noted.

Figure 1 is a graphic illustration of the combined RLIN steps presented in table 2 and clearly shows that the type of RLIN record being established has a definite effect on the conversion time (and therefore on the cost) of the record. Differences be-

Table 1

Step	Staff Level	St. Time (Minutes)	Cost (Dollars)
SL Categorization	Hourly	0.0235	0.0013
Photocopy on Form	Hourly	0.0177	0.0010
Sort Forms	Clerical	0.0586	0.0074
*Search Public Catalog	Hourly	1.7537	0.0979
*Search Authority File	Clerical	1.2123	0.1532
*Tag Record	Clerical	1.5419	0.1948
*Resolve Problems	Clerical	1.9656	0.2484
RLIN Work	Clerical	3.6369	0.4596
RLIN ID on SL Card	Clerical	0.5297	0.0669

Table 2

Rec. Type	Searching	Fixed Fields	Var. Fields	Holdings	Problems	Writing ID
Create	1.06	0.34	4.36	0.38	0.38	0.13
Create *	1.18	0.46	2.91	0.45	0.30	0.14
Derive	0.68	0.25	1.46	0.52	0.32	0.14

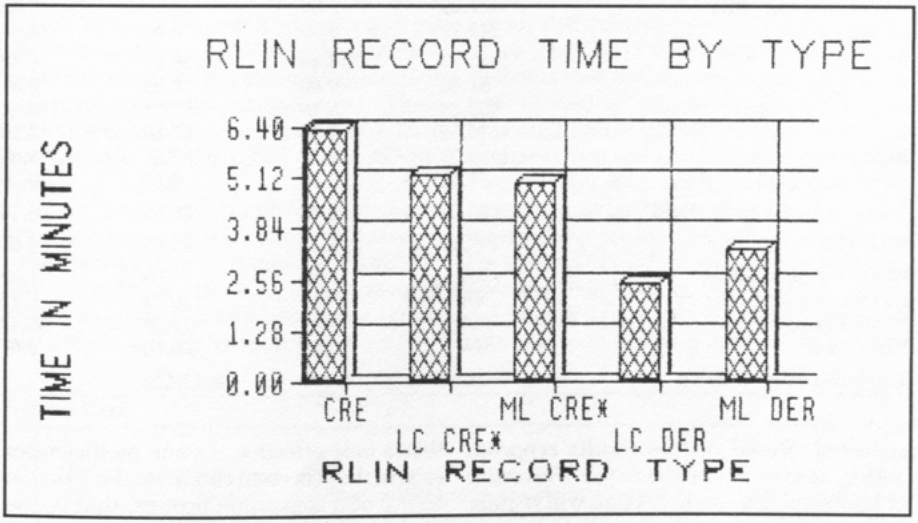


Fig. 1. RLIN Record Time by Type.

tween member library and Library of Congress copy are also shown. Creating a new record on RLIN, from searching it through keying in fixed and variable fields, takes more than twice as long as deriving a local record from an exact match. Changing an edition record to suit the copy in hand is also more time-consuming than deriving a record.

These times become more meaningful when the sample population is analyzed. The types of records seen during the study were:

20.2% Create
*10.9% Create **
69.0% Derive

Even though *Create* records represented only 20.2 percent of the total population, they represented 43 percent of the RLIN time. *Create* * and *Derive* represented 35.2 percent and 21.8 percent, respectively, of the RLIN time.

Step 10 (total combination of records and combination of pre-RLIN work) was added to the study because of the perceived variation in the performance of steps from record to record. The analysis of this data, combined with the RLIN results listed above, allowed percentages of steps (and therefore times and costs) to be drawn up for a population of one thousand items. Using this population then allowed us to describe an average record that would account for the record variations in an accurate way. Table 3 provides time and cost figures for each step in the complete conversion process for a one thousand–record average based on our sample. On the average it takes *7.82* minutes and costs *$2.12* to convert one record into machine-readable form. The cost breaks down as follows:[4]

Staff cost:	$0.93
RLIN telecommunications and maintenance cost:	0.17
RLIN communications and CPU cost:	1.02

ANALYSIS

Clearly the percentage of hits directly affects the amount of data entry that will be

Table 3.

Step Performed:	Time Minutes	Percent Time	Cost Dollars	Percent Cost
SL Categorization	24.00	0.31	1.34	0.06
Photocopy Cards on Forms	177.00	2.26	9.88	0.47
Sort Workforms	586.00	7.49	74.05	3.50
Tag Records	1401.59	17.92	177.08	8.36
Search Public Catalog	859.31	10.98	47.87	2.26
Search Authority File	147.90	1.89	18.70	0.88
Problem Solve	241.77	3.09	30.55	1.44
RLIN *Create*	1133.73	14.49	136.57	6.45
RLIN LC *Create* *	51.80	0.66	6.55	0.31
RLIN ML *Create* *	403.38	5.16	50.98	2.41
RLIN LC *Derive*	536.55	6.86	67.80	3.20
RLIN ML *Derive*	1484.05	18.97	187.54	8.85
RLIN Cancelled	77.52	0.99	9.80	0.46
Write RLIN ID on SL Card	530.00	6.78	66.98	3.16
Supervisor	168.49	2.15	43.63	2.06
RLIN Terminal Maint.	NA	NA	39.88	1.88
RLIN Communications	NA	NA	129.74	6.12
RLIN Transactions	NA	NA	814.02	38.42
RLIN CPU	NA	NA	205.50	9.67
Totals [Minutes/Dollars]:	7823.09		$2118.56	

required. Based on the results reported above, conversion of the entire University of Michigan Library collection will require a substantial amount of direct data entry in the form of (1) creating records for which there are no matches in the resource file and (2) modifying records copied from the resource file (the exact effort that would be required would depend on the specific resource file used). While the specific costs depend on a number of factors it is fair to say that conversion of the entire collection will be costly in the way described above, so costly that it is unlikely the collection will be converted entirely in this way.

Several strategies exist for reducing the costs associated with retrospective conversion. First, one must select a resource file that maximizes the number of record matches. Second, it is possible to elect to convert only a portion of the collection. However, research has shown that users want online catalogs to contain all available records and that they will not use a card catalog when an online catalog is available.[5] It appears, therefore, that users will find unacceptable those online catalogs that contain only a small portion of the collection. Third, efforts can be made to streamline the conversion *process* itself. But this is only effective if some inefficiencies exist in the process to eliminate. Fourth, the choice of a conversion process, that is, use of a utility, batch processing through a commercial vendor, etc., can influence the ultimate cost.

A fifth alternative, which has been occasionally mentioned (and even implemented, primarily for *circulation*), is a reduction in the *content* of the record converted.[6] This would be a possible option for those items that did not match a record in the resource file. However this reduction in record content is usually seen as being less than the ideal—a complete "catalog" record containing all the details of the original record plus any additional details needed to provide full MARC format. The "complete" record may also be upgraded to current cataloging standards. A brief record is frequently seen as temporary, filling a practical need, and is not viewed as an alternative to the "complete" record.

The catalog record has served as both a "finding tool" (call number, subject, and other entries that serve as access points in a catalog, either card or online) and a "reference tool" (pagination, illustrations, bibliography, contents notes, and the like). While it is appropriate for an online catalog

to serve in the same way, the economics of retrospective conversion may not leave the choice open for records that cannot be converted using resource files. If it is not possible to create "full" MARC records for records that must be originally converted, we must ask ourselves: Is it better to provide access to a record composed of a subset of MARC that fulfills users' and staff needs for information retrieval? We believe the answer to that question is yes and suggest that such records be called *access records*. These records would follow the MARC structure but would employ a subset of the MARC fields.

An important question is What elements are needed in a record to ensure full and accurate retrieval by users and staff? The definition of the content of access records would be guided by two principles: (1) retrievability and (2) recognition.

If an access point retrieved a record so spare as to be unrecognizable, the retrieval would be meaningless. Thus an access record must be composed of both access elements, such as authors, titles, and subjects, and descriptive elements such as imprint. In addition, records might accommodate differences in information retrieval needs among disciplines.

It should be possible therefore to define elements that will appear in all access records as well as suggest optional elements that might vary by discipline. Mandatory elements (variable fields only) would be 1xx, 2xx, 6xx, and 7xx fields. Optional elements would be 300, and 5xx fields. Serial information for monographs contained in the 4xx and 8xx fields, would have to be adjusted to individual institutions' needs.

In figure 2 the original cataloging record contains four access points (author, title,

Catalog Card:

DC
198
.D36
S88

Stouff, Louis, 1859–
Le lieutenant général Delort, d'après ses archives et les archives du Ministère de la guerre 1792–1815, par L. Stouff . . . Avec 4 planches et 5 croquis. Paris, Nancy, Berger-Levrault & cie, 1906.
3 p.l.,128,177 p. 2 pl., 2 port. (incl. front.) 5 fold. maps. 25cm.
"Sources": [1.ptie.] p.[1]–4: "Oeuvres du général Delort'; [2.ptie.] p.[161]–166.
CONTENTS.—[1.ptie.] Essai sur le lieutenant général baron Delort.—[2.ptie.] Papiers du Lieutenant général baron Delort (1792–1815)

1. Delort, Jacques Antoine Adrien, baron, 1773–1846.
2. France—History, Military—1789–1815. [Full name: Marie Pierre Xavier Louis Stouff]

MiU36-1167

O

Access Record (variable fields only):

100 10 Stouff, Louis, $d1859–
245 13 Le lieutenant général Delort, d'après ses archives et
et les archives du Ministère de la guerre 1792–1815
260 0 $c1906
600 10 Delort, Jacques Antoine Adrian, $cbaron, $d1773–1846
651 0 France $xHistory, Military $y1789–1815

MARC Elements (variable fields) *Not* Included:

245 13 $cpar L. Stouff. Avec 4 planches et 5 croquis.
260 0 Paris, $aNancy, $bBerger-Levrault & cie.,
300 3 p. 1; 128; 177 p.$b2 pl., 2 port. (incl. front.) 5 fold. maps. $c25 cm.
504 "Sources": [1.ptie.] p. [1]–4; "Oeuvres du général Delort": [2.ptie.] p. [161]–166.
505 0 [1.ptie.] Essaie sur le lieutenant général baron Delort.—[2.ptie.] Papiers du Lieutenant général baron Delort (1792–1815)

Fig. 2.

and two subjects); imprint; somewhat complicated pagination; information concerning plates, portraits, and maps; source information; and contents. The created *access record* contains the imprint and all four access points but excludes all other information.

The original cataloging record for figure 3 contained six access points (author, two titles, two subjects, and one personal name–added entry) and considerable information regarding authorship, publishing, imprint date, pagination, printing, contents, and name information concerning the author. The *access record* includes the imprint and all access points except the title-added entry but excludes all additional information. Title-added entry would be retrievable through Boolean searching. As the example demonstrates, the full MARC record for the variable fields would require fifteen more lines of keying.

Figure 4 is much like the other two above, with access points still the same. In this particular case, however, the collation information, particularly the presence of il-

Catalog Card:

H
62
.I59

Institut scientifique de recherches économiques et sociales, Paris.

. . . Cinq conférences sur la méthode dans les recherches économiques, par Lionel Robbins . . . E.F. Wagemann . . . [et autres] Paris, Librairie du Recueil Sirey, 1938.

3 p. l., [9]–106p., 1 l. 19cm. (Its Mémoires et enquêtes. VI) Printed in Belgium.

CONTENTS:—Robbins, Lionel. Les méthodes d'observation économique et les problèmes de la prévision en matière économique.—Wagemann, E. F. Organisation et méthodes de travail de l'Institut allemand pour l'etude de la conjoncture.—Dupriez, Léon. Enseignement et recherches économiques à l'Université de Louvain.—Vandellòs i Sola, J.A. État actuel des recherches économiques en Espagne.—Verrijn Stuart, G.M. Les études économiques aux Pays-Bas.

1. Economics—Study and teaching. 2. Research. I. Robbins, Lionel Charles, 1836– II. Title. III. Title: La méthode dans les recherches économiques.

Library of Congress H62.I65 42-7068
330.72

O

Access Record:

110 20 Institut scientifique de recherches économiques et sociales, Paris
245 10 Cinq conferences sur la méthode dans les recherches économiques
260 0 $c1938
650 0 Economics. $xStudy and teaching
650 0 Research
700 10 Robbins, Lionel Charles, $d1898–

MARC Elements (variable fields) *Not* Included:

245 10 $cpar Lionel Robbins . . . E.F. Wagemann . . . [et autres]
260 0 Paris, $bLibrairie du Recueil Sirey,
300 3 p. l., [9]–106 p.,1 l.,$c19 cm.
490 0 Its Mémoires et enquetes. $vVI
500 Printed in Belgium
505 0 Robbins, Lionel. Les méthodes d'observation économique et les problèmes de la prévision en matière économique.—Wagemann, E.G. Organisation et méthodes de travail de l'Institut allemand pour l'étude de la conjoncture.—Dupriex, Léon. Enseignement et recherches économiques à l'Université de Louvain.—Vandellòs i Sola, J.A. État actuel des recherches économiques en Espagne.—Verrijn Stuart, G.M. Les études économiques aux Pays-Bas.
740 31 La méthode dans les recherches économiques.

Fig. 3.

Catalog Card:

FINE
ARTS

ND
212
.P23

Paris. Musée national d'art moderne.
Jackson Pollock et la nouvelle peinture américaine. [Exposition] 16 janvier—15 février 1959.
[Catalogue] Paris, éditions des Musées nationaux [1959] [48] p. plates. 25 cm.
"L'International Council at The Museum of Modern Art a l'honneur de présenter réunies â Paris les deux expositions 'Jackson Pollock: 1912–1956' et 'La nouvelle peinture américaine' qui furent toutes deux organisées . . . par le Programme Internationale du Museum of Modern Art de New York."
1. Pollock, Jackson, 1912–1956. 2. Paintings, American—Exhibitions. 3. Painters, American—Exhibitions. I. New York. Museum of Modern Art. International Program.

O

MiU

Access Record:

110 10 Paris. $bMusée national d'art moderne.
245 10 Jackson Pollock et la nouvelle peinture américaine. $bExposition 16 janvier–15 février 1959. [Catalogue]
260 0 $c 1959
300 [48] p.$bplates$c25 cm.
600 10 Pollock, Jackson, $d1912–1956.
650 0 Paintings, American. $xExhibitions.
650 0 Painters, American. $xExhibitions
710 10 New York. $bMuseum of Modern Art. $bInternational Program.

MARC Elements (variable fields) Not Included in Access Record:

260 0 Paris, $bEditions des musée nationaux
500 "L'International Council at the Museum of Modern Art a l'honner de présenter réunies â Paris les deux expositions 'Jackson Pollock: 1912–1956' et 'La nouvelle peinture américaine qui furent toutes deux organisées . . . par le Programme Internationale du Museum of Modern Art de New York."

Fig. 4.

lustrations, is of more importance than similar elements in the other examples and has been included. This element should be a selected optional element for all fine arts records with that field.

It is important to recognize that online searching is potentially more efficient than card catalog searching because of the ability to use key words, truncation, Boolean searches, and stop lists. Such searching capacity may completely eliminate the need for some aspects of full cataloging, such as alternate titles or reciprocal subject headings.

CONCLUSIONS

One must remember that the original impetus for looking at records in anything but the traditional way was to reduce cost. There is an underlying assumption that the shorter record is faster to convert because of the reduction in the number of fields. It is also an assumption that some of the elements not included in the mandatory list involve a disproportionately large amount of time to convert in both the offline and online work. Many variations in the *access* record might be time-consuming to monitor, particularly if they varied from record to record. The dilemma is that no brief record format fills the needs of all disciplines.

While the advantages of a more complete record are understood, the economics of *access* records constitutes a strong argu-

ment for their use. While the use of vendor- and utility-supplied resource files does much to reduce the cost of retrospective conversion, the fact remains that a substantial portion of the collections in research libraries is not represented in the resource files. A recent study of six libraries—Michigan, UC Berkeley, Columbia, Cornell, Pennsylvania, and Yale—indicated that 20 percent of the sample population was held by only one library and that more than one-third was held by two or fewer libraries.[7] The same study found that one-third of the sample was not represented in the OCLC database, and the figure for RLIN was 50 percent. Clearly research libraries will have to key a sizable percentage of the records they convert. By reducing the time it takes to key, conversion costs can be reduced, and one way to reduce keying time is using access records. We believe the concept of access records for original retrospective conversion deserves further debate and discussion.

ACKNOWLEDGMENTS

Our appreciation to Mary Alice Ball for assistance in collection of the data, to Robert "Mac" McDonald and Chris Wilcox for assistance in the formatting and testing of sample records, and to Lynn Marko and Robert Houbek for discussion and review of much of our thinking.

REFERENCES AND NOTES

1. Stephen H. Peters and Douglas J. Butler. "A Cost Model for Retrospective Conversion Alternatives," *Library Resources & Technical Services* 28:149–62 (1984).
2. Ibid, p.154.
3. There are three types of RLIN records: *Create* (when no online record exists), *Create* * (when an online record for an edition exists), and *Derive* (when a matching record exists).
4. RLIN costs are based on April 1985 billings and production figures and include calculations for peak and offpeak rates and actual terminal use and maintenance.
5. Ferguson Matthews and others, *Using Online Catalogs* (New York: Neal-Schuman 1983), p.174–75.
6. Michigan's Brief Record Cataloging, produced for some new materials, resembles the access record. It takes approximately four minutes to search and key in a Brief Cataloging record. This compares favorably with the time noted in the Time Study for a *Create* record, representing a reduction of about 40 percent in time for inputting alone.
7. Robert Houbeck, Jr. and others. Unpublished collection overlap study, p.3. ■■

Structured Analysis and the Data Flow Diagram: Tools for Library Analysis

David H. Carlson

The meanings of the terms *system* and *systems analysis* vary widely. In management, the terms have associations of organizational and management structure; in science, they mean the study of scientific paradigms and scientific communication; in engineering, the terms often mean the coordinated management of a complex and lengthy technical project. "Systems analysis is a familiar expression but it is also poorly understood."[1] That statement of five years ago in a lead article from *Choice* is as true today as it was when written. Simply put, systems analysis is the study of systems, whether these systems are corporations, scientific communities, or technical projects.

Specialized computer dictionaries typically define systems analysis as, "the examination of an activity, procedure, method, technique, or business to determine what must be accomplished and the best method of accomplishing the necessary operations."[2] Even though computer dictionaries make no mention of the computer in their definitions of systems analysis, an increasingly popular perception of systems analysis is that it is the specific process of applying the computer to a problem.

There are good reasons for the popular association of the computer with systems analysis. The successful application of the computer requires a thorough understanding of the manual system the machine is intended to replace or supplement. It is the goal of systems analysis to provide such understanding. Another reason for the association is that in the past twenty years, the solutions proposed as a result of systems analysis have often been computer based. As the price of computer hardware decreases and the quality of software improves, this trend will surely continue and expand. Finally, as an activity that requires thorough understanding, computer programming itself benefits from systems analysis.

TOOLS OF ANALYSIS

Many new tools have been developed to aid the systems analysis process. In developing these tools, certain characteristics have been found to be important for effectiveness. One such characteristic is assistance in problem solving. Tools that are a problem-solving aid do more than document the system. They help the systems analyst understand the present system, allow for easy identification of problem areas, and, ideally, suggest effective solutions.

A second characteristic of a useful tool of analysis is that it must also be an effective communications device. The need for good communications is critical to successful analysis. The analyst must communicate to management the strengths and weaknesses of the present system, the basic features of a proposed solution, possible alternative solutions, and any problems the solutions do not address. Good communication is also essential for the analyst in working with staff. Those who work with the system day in and day out are often the most important source of information on the present system, and they are the ultimate authority on whether the analysis of the present system is accurate. Staff may be the critical determining factor in the success of a future system as well.

David H. Carlson is systems analyst, University of Rhode Island, Kingston.

The dialogue between analyst, staff, and management is enhanced and facilitated when the analyst can use a tool that lends itself to communication and discussion with these groups. "The most successful projects are truly joint efforts in which both analysts and users make significant contributions."[3] The key to achieving a joint effort is through communication.

The traditional narrative report is the most common and popular vehicle to communicate the results and recommendations of an analysis. As a communications tool, however, the narrative report has limitations. Its word-by-word, sentence-by-sentence structure demands a perusal and time commitment that many, particularly administrators, are not wiling to give. But if it is the only tangible result, the narrative may do more than communicate results; by default, it may become a primary tool in conducting the analysis as well.

The narrative may have shortcomings as a communications device but when it becomes a tool of analysis as well, it is woefully inadequate. The analyst needs other tools that will graphically demonstrate system problems and inadequacies, that will uncover inconsistencies in policy, and that will complement the narrative in communicating results. Many tools have been developed as an alternative or supplement to the narrative approach. Most employ some set of graphic symbols. Some of the more popular methods are listed below, followed by a brief discussion and explanation.

1. PERT: Program evaluation and review technique. PERT was originally developed by the U.S. Navy for development of the Polaris Ballistic Missile. Its use for Polaris is indicative of the kind of project in which PERT is most useful—the management of a complex, technical project that requires the coordination of many separate activities within a specific time frame.

The primary graphic symbol of PERT is a network. The network is composed of nodes or circles connected by lines. The nodes represent the beginning or end of an activity; in the terminology of PERT, nodes are referred to as events. The lines represent activities or tasks necessary for completion of the project.

Each activity is assigned time estimates. Three estimates are usually given: optimistic time, most likely time, and pessimistic time. With this information, a PERT chart will identify the critical path for the project manager. The critical path determines the minimum time for completion of the project. The critical path identifies those activities that are critical to the project's successful and timely completion. For this reason, PERT is frequently referred to as CPM, critical path method.

A classic and detailed examination of PERT was done by Archibald and Villoria.[4] A less detailed, tutorial approach is provided by Evarts.[5] While there are brief explanations of PERT in library management texts[6] there are few published applications of PERT in libraries.

2. Gantt Charts. Gantt charts were a primary factor in the development of PERT. The basic concept of Gantt charts was developed by Henry L. Gantt during World War I. In Gantt charts, activities are represented as a series of horizontal bars with the beginning, duration, and end of the activity represented by the bar as plotted against a time scale.

With the beginning of World War II and the atomic age, however, technical projects became more complex and difficult to manage than ever before. One of the complexities that proved most difficult to deal with was the sophisticated interdependence between activities. An example of interdependence in a technical project is the need to have activity X completed after activity Y but before activity Z, without which Y would not work. In a complex technical project there may be hundreds of such requirements. They must be effectively coordinated while work on several interrelated activities occurs simultaneously. The strict sequential Gantt charts do not account well for interdependence and the coordination it requires.

3. Organizational Charts. Organizational charts are one of the most familiar graphic tools of analysis. They depict the structure and lines of authority in an organization.

The graphic elements of an organizational chart are rectangles with connecting lines. The rectangles represent a position or department in the organization. Organizational charts are hierarchical. The placement of a position or department on the

chart represents the responsibility and lines of authority relative to other departments and positions in the organization.

4. Decision Tables. Decision tables are used in analysis for documentation of complex policies. On the left side of the decision table the conditions and possible actions of the policy are listed. The table is completed by indicating the appropriate action for each set of conditions.

For example, if its complexity called for it, a decision table could be made of a library's circulation policy. On the top left side of the table the analyst would list all the conditions that determine whether a patron may check a book out of the library. Residency might be one condition; type of material (e.g., a reference book) might be another. Below these conditions would be listed all possible actions. In this example, alternatives might include whether to permit circulation for a limited period of time or conditional circulation, pending approval by an individual or department.

A primary advantage of decision tables is that their structure forces an assessment of each possible set of conditions. If all conditions and actions are listed, the decision table ensures that all resulting possibilities will be represented. Once each possibility is fully identified by the table, the analyst indicates the appropriate action for each situation in the body of the table. This characteristic of decision tables is a good example of a tool that is an aid in problem solving. The decision table does more than document complex policies; it demonstrates to the analyst where discrepancies or vagaries in policy may exist. There is an excellent chapter on decision tables in Weinberg[7]; other texts give the tool full review and discussion.[8,9]

5. Flowcharts. Flowcharts are another well-known tool of analysis. They enjoy a comparatively long history and come complete with an official ANSI standard.[10] One reason the flowchart is a useful tool of analysis is that it can represent a wide range of processes. It may be used for the analysis of a system, such as a library, as well as the logic of a computer program. The limitations of flowcharts are discussed in detail below.

6. HIPO: Hierarchy plus input-process-output. HIPO is a complete documentation technique that refers to the use of three graphic tools: a visual table of contents (VTOC), overview HIPO diagrams, and detail HIPO diagrams. The basic graphic elements of the detail HIPO diagram are three large rectangles arranged side by side on a sheet of paper. The left rectangle contains input tasks, the center rectangle processes or steps, and the right rectangle the resulting output. Interaction between these three basic functions is represented with arrows from one rectangle to the other. The type of material that appears in the three rectangles varies, ranging from simple narrative statements (e.g., "Generate Exception Report") to a small set of graphic symbols. While HIPO symbols do not enjoy an official ANSI standard as do flowcharts, they are nonetheless relatively stable.

The HIPO technique tends to be a very detailed one. The detail is offset, however, by the more general approach of the other two tools that are part of a complete HIPO approach to analysis, the overview diagrams, and VTOC. The overview diagrams are HIPO charts drawn specifically to provide an overview rather than detailed analysis. The VTOC is graphically similar to an organizational chart except that it represents the hierarchical and structural approach of a computer program. Katzan provides a comprehensive explanation of the HIPO technique.[11]

COMMON CHARACTERISTICS

It is useful to consider briefly some similarities and differences among these various techniques. One common characteristic is the use of graphic elements. There are two reasons for the emphasis on graphic tools: first, graphic tools are an aid in communication, and second, a graphic representation is more representative of a system than a narrative.

> An additional advantage of pictures is that they are two-dimensional, whereas text is linear—that is, one-dimensional. The patterns within a system, and its overall shape, are at least two-dimensional. Therefore, a graphic tool is more 'readable' than a verbal one, can be more accurately interpreted by the brain and is a more realistic description of a system.[12]

It is also interesting to note that while there is little in common between the individual graphic symbols of these six tools, each ap-

proach limits itself to a small, limited set of symbols. A small set of graphic elements keeps the tool simple and readily understood. Communication is enhanced when the tool is unobtrusive and easily explained.

The above six methods are tools of the analyst. However, one primary difference among them is that the likelihood of their use depends extensively on the particular area of application. For example, GANTT and organizational charts are used predominantly in business applications; PERT is used extensively in technical projects; flowcharts and VTOCs are used heavily in data processing, and so on. This is true because some methods are better than others at analyzing certain types of systems and problems.

Finally, each of the six tools has limitations and shortcomings. Simplicity and ease of understanding are critical features, but the simplest approaches tend to be limited to specific, specialized applications. The organization chart and VTOC are good examples. Graphically, these tools could not be simpler or easier to understand, but their effectiveness is limited to the hierarchical structure of an organization or the similar hierarchical structure of a computer program. The flowchart is an example of the other extreme. It is an extraordinarily flexible tool with applications in many different situations and problems, but it is not as graphically simple or as easy to understand as the organization chart or VTOC.

None of these six tools is in competition with another. Just as a carpenter has a large variety of tools with which to build, so the analyst may choose from these and other tools to accomplish a successful analysis. The variety of approaches and the strengths and limitations of each tool are more appropriately characterized as advantages that the analyst can exploit and use. What is important is that each tool is used to its full potential in its application to a specific problem.

LIBRARY APPLICATIONS OF ANALYSIS

A specialized computer dictionary defines *system* as "an assembly of components united by some form of regulated interaction to form an organized whole."[13] Libraries clearly fit this definition. The components of acquisitions, cataloging, reference, circulation, etc., interact and work together to provide the library's community of users with information and materials. Libraries are typically part of larger systems with more comprehensive goals, such as a university, a town government, or a business. Since the use of a tool varies on the application and since some tools are better than others at analyzing certain types of problems and systems, are there tools that are well suited to analyzing a library as a system?

There have been many applications of flowcharts in libraries.[14] From the author's experience with library schools, library education appears to endorse flowcharts as the tool of choice, as they are frequently taught in one or more information science classes in the library school curriculum.

Some of the positive elements of flowcharts were mentioned earlier: extraordinary flexibility, simple and standarized graphic symbols, and wide acceptance. However, as a tool for systems analysis, flowcharts have several critical problems. The most significant is the extensive level of detail that flowcharts require. Every decision within the system, every branch, every process, every file must be represented, or the flowchart appears incomplete and unrepresentative. As a method for gaining a thorough understanding of the system, this in-depth approach is useful. However, the flowchart's detail prohibits the kind of simplification and overview that often lead to broad-based understanding and insight. It is difficult for the analyst to identify problems or for the flowcharts themselves to suggest solutions; the detail is overwhelming and constrictive.

Detail is not only required of the analyst for flowchart development but an understanding of it is also required of those who will view the chart. As a tool for communication with management or staff, the flowchart is not effective. Administrators are simply not interested in the comprehensive analysis of a flowchart, and staff are intimidated by it.

To overcome these problems some authors recommend "overview flowcharts."

These use the same flowchart symbols but are drawn specifically with the intent of generalizing and providing an overview. This solution does not solve the problem. The flowchart's detail, its meticulous step-by-step analysis, is not simply a reflection of the analyst's goal in drawing the chart but is also a reflection of the overall approach. The strictly linear graphic presentation and the selection of flowchart symbols lend themselves well to detail but become major obstacles for a more general, broad-based analysis.

Finally, the flowchart fails as an aid for developing the logic of a computer program. The detail it requires is so exacting that most programmers, when required to use flowcharts, use them only after the program is running, not to develop it.[15]

STRUCTURED SYSTEMS ANALYSIS

This paper focuses on a relatively new tool, the data flow diagram (DFD) as a tool for the analysis of a library. The DFD is one element of a broader approach to analysis usually termed *structured analysis*. Before examining the DFD and its application in the library, the broader context of structured analysis in which the DFD was developed should be considered.

Since 1974 and the publication of a paper by Larry Constantine in the *IBM Systems Journal*[16] there has been much discussion and attention in the field of data processing to concepts and terms that begin with the word *structured*: structured analysis, structured design, structured programming, and structured charts. These terms refer to some element of a complete systems analysis. Each has unique tools and concepts but differs in timing, specific objectives, and other areas. They share some fundamental approaches and concepts, nonetheless, and are all associated with the structured movement in data processing.

DeMarco is another author important in the development of the structured approach. His book, *Structured Analysis and System Specification*,[17] is one of the earliest texts on structured analysis, and it includes detailed information on the development and use of a DFD.

Constantine, DeMarco, and other like-minded authors were employed by the same consulting firm, Yourdon. Yourdon specializes in structured systems analysis to organizations and data processing managers. Besides consulting, Yourdon has a healthy publishing sideline. The list of Yourdon books is devoted almost solely to titles on some aspect of the structured approach, and several Yourdon titles are considered classic texts. One of the reasons the structured approach has received so much attention is that it has proven to be a successful method of analysis and design.

The approach to DFDs, first outlined by Constantine, developed more fully by DeMarco, and proselytized by Yourdon, was modified by Chris Gane and Trish Sarson in 1979. Their refinement was published in a book entitled *Structured Systems Analysis: Tools and Techniques*[18] in 1979. While the Yourdon, Gane, and Sarson approaches differ in several areas, their basic approach to analysis is very similar. The Gane and Sarson refinement of DFDs and their application in libraries is discussed below.

THE DATA FLOW DIAGRAM

The data flow diagram was developed for the analyst as an aid in problem solving and as a vehicle for effective communication. The DFD is also meant to be both simple and flexible in order to represent a wide range of systems and problems. The DFD employs just four different graphic symbols: a rounded rectangle; a heavy-sided square; a flat, open-ended rectangle; and arrows connecting these symbols. Figure 1 is a DFD of one library's analysis of a patron's search for an issue of a serial. It will serve as an example in the following discussion.

The first and most important symbol of a DFD is the rounded rectangle. It represents steps or processes. Two examples of processes from figure 1 are "Check Shelf for Volume" and "Initiate Recall of Volume." Processes are the activities or steps a patron or staff member performs. Numbering the processes in the upper portion of the symbol is done generally from left to right along a typical or common path. The numbers are not especially important, however, as they serve primarily for identification.

The second symbol, a heavy-sided square, represents areas that, while part of

the overall system, are outside the scope and focus of the DFD. These symbols are referred to as external entities or sinks. They may be thought of as black boxes. "Where the system we are considering accepts data from another system or provides data to it, that other system is an external entity."[19] Some data go into an entity or come out, but how that is accomplished is not a concern of the analysis.

An example of a sink from figure 1 is interlibrary loan. As a library function, interlibrary loan is clearly an important part of meeting patron needs for periodical articles. However, this DFD treats ILL as an external entity. ILL is part of the larger perspective (a patron's search for a journal issue), but the focus of this particular DFD is to diagram the search for a journal strictly from the perspective of the patron, not the library. From this perspective, ILL is very much like a black box. Something goes in, an ILL request, and something comes out, a photocopy of the article, but how it is all done is not a concern. It is therefore represented as a sink on the DFD.

The third symbol, a flat, open-ended rectangle, represents files. The files represented by the rectangles may be in numerous formats. They may be locally maintained card files (a Kardex), standard works (*Books in Print*), or even external computer databases (OCLC). What is common to these is that they are all files, places where data is stored and maintained.

The final symbol is the arrow that connects the symbols. The arrows represent the flow of information or data through the system.

In some cases, a symbol appears more than once. Interlibrary loan, for example, appears five times in figure 1. This is perfectly acceptable and is done to make the charts appear less complex with a minimum of crisscrossing lines.

ADVANTAGES OF THE DFD

The DFD shares graphic simplicity with the other tools discussed above. Its limited number of symbols and basic concepts are quickly explained and readily understood. Primarily for this reason, it is an effective and useful communications device.

One primary difference between the DFD and many other tools of analysis is that the DFD incorporates concepts and principles of structured analysis. For example, an important tenet of structured analysis is that the details of implementation are part of the later stage of system design and an inappropriate concern of analysis. The structured movement maintains that analysis is a logical process above all else and that the physical details of how an analysis might eventually be implemented are best considered once the analysis is complete and the process of design begins.

In support of this principle, the DFD has no symbols for different hardware devices. As mentioned earlier, the flat rectangular symbol of the DFD represents all types of files, be they card files or online databases. The rounded rectangle represents all types of processes. The DFD is not concerned with how the process is carried out. This lack of symbols for hardware devices can be contrasted with the flowchart, for example, which has separate symbols for devices such as printers, disks, tapes, and others.

Another principle of structured analysis is that systems be approached "top-down." This method maintains that analysis and design begin at the top, most general conceptual level and then, piece by piece, work down to lower, more detailed levels. While there are several beneficial results from a top-down approach, one of the most important is that the analysis can be broken down into processes that are distinct, separate, and manageable.

In his book, *Mindstorms*, Seymour Papert discusses the importance of this characteristic for children learning the computer programming language LOGO.[20] Papert borrows a term coined by one of the children in his study, *mind-size bites*. The term is especially descriptive. People, be they children, adults, analysts, or library directors, can accept only so much information before it becomes overwhelming. Because of this, the goal of a top-down approach is processes that are mind-size bites.

Figure 1, for example, would never stand entirely on its own. In a full analysis it would be supplemented by a series of lower-level processes expanding the higher-level ones. Thus, process fourteen, "Locate

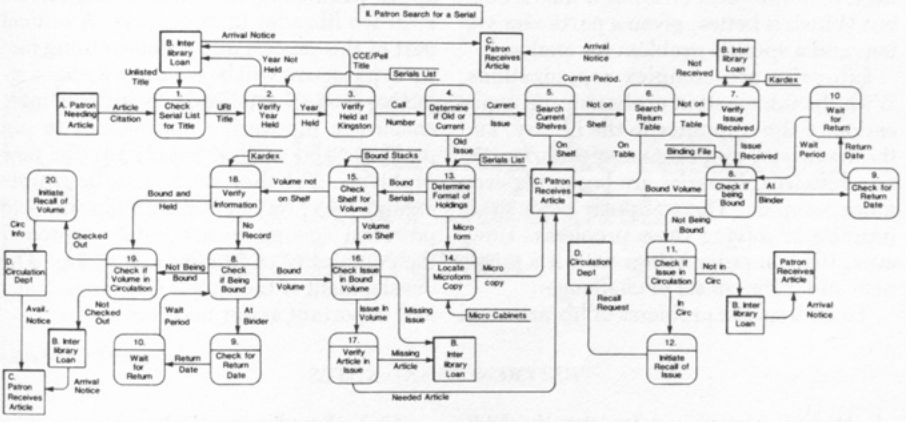


Fig. 1.

Microform Copy," might well be expanded at a lower level to show the processes, files, and external entities necessary to "Locate Microform Copy." Depending on the complexity of this expansion, further refinement might be necessary. The refinement is repeated until all necessary details are represented.

Independent and manageable processes are important for effective analysis, but they have a second advantage as well—improved communications. Tools that incorporate a top-down approach, such as the DFD, begin at the top, most general level of conceptualization with details "hidden" at lower levels. In this way when the results of the analysis are presented to other groups, details are approached as needed and as questions arise—that is, as aids to communication—and *not* as the first obstacles to overcome in understanding the analysis.

However, for the analyst there is a conflict between the simplification necessary for communication and understanding and the detail required by the machine. The computer is not limited in the amount of detail it can process. Indeed, the precise working through of every detail, every branch, every possible state and condition is the essence of computer programming. One of the greatest challenges of analysis is to communicate both with people, who can accept limited information and detail, and with the machine, which requires precise, detailed instructions for every possible situation and process.

The DFD resolves the conflict between communication and detail by transforming the detail into a vehicle for communication. For example, the expansion discussed above for process fourteen, "Locate Microform Copy," adds the detail necessary for a computer-based design, but these added details do not change process fourteen in any way; it is still "Locate Microform Copy." For those for whom this general description is enough, the detail is hidden; for the analyst, for those who work in microforms, or for administrators concerned about a critical or expensive procedure, the detail is there to be discovered and explored. In this way, the detail becomes part of the communications process, not an obstacle to it.

CONCLUSION

In this paper, the advantages of the DFD have been explained to bring attention to a tool that has great potential for library analysis but that has seen very little application. It was noted earlier, and it is good to be reminded here, that tools such as the flowchart and DFD are not in competition with each other. DFDs, flowcharts, and other tools have particular strengths and weaknesses that can be used and exploited to advantage by the analyst. In evaluating

these tools the question is not Which is best? but Which is better, given a particular system and a specific problem for analysis?

Libraries are complex organizations. With the information explosion, the application of the computer in the library, and the increasing importance of information as a resource, libraries are becoming even more complex. The computer holds great promise in solving these problems. However, the computer brings with it a whole new and different set of challenges.

To address the problems of libraries and apply promising new technologies, we must see libraries in new ways. A critical part of this process and a determining factor in its success will be whether we have effective tools of analysis. These tools must encourage innovative approaches to old problems and suggest fresh ideas for new problems. Yet they must also be simple enough to enhance communication and powerful enough to accurately represent the complexity of library processing. The development of these tools is as challenging and important as our need for them.

REFERENCES AND NOTES

1. Michael McGill and Jennifer Huitfeldt, "Systems Analysis," *Choice* 15: 1623–32 (Feb., 1979).
2. Charles J. Sippl and Roger J. Sippl, *Computer Dictionary*, 3d ed. (Indianapolis, Ind.: Sams, 1980), p.549.
3. Victor Weinberg, *Structured Analysis* (Englewood Cliffs, N.J.: Prentice-Hall, 1980), p.4.
4. Russell D. Archibald, *Network-Based Management Systems (PERT/CPM)* (New York, N.Y.: Wiley, 1967).
5. Harry F. Evarts, *Introduction to PERT* (Boston: Allyn and Bacon, 1964).
6. Robert D. Stueart and John Taylor Eastlick, *Library Management*, 2d ed. (Littleton, Colo.: Libraries Unlimited, 1981), p.161–2.
7. Weinberg, *Structured Analysis*, p.113–24.
8. Thomas R. Gildersleeve, *Decision Tables and Their Practical Application in Data Processing* (Englewood Cliffs, N.J.: Prentice-Hall, 1970). This text is primarily a compilation of decision table examples.
9. Solomon L. Pollack, Harry T. Hicks, and William J. Harrison, *Decision Tables—Theory and Practice* (New York, N.Y.: Wiley, 1971).
10. N. Chapin, "Full Report of the Flowchart Committee on ANS Standard X3.5-1970," *Sigplan Notices* 14:16–27 (Mar., 1979).
11. Henry Katzan, *Systems Design and Documentation* (New York, N.Y.: Van Nostrand Reinhold, 1976).
12. Meijer Page-Jones, *The Practical Guide to Structured Systems Design*, (New York, N.Y.: Yourdon Pr., 1980), p.24.
13. Sippl and Sippl, *Computer Dictionary*, p.544.
14. Many articles could be cited. One of the earliest articles on the application of flowcharts in libraries was C. D. Gull, "Logical Flow Charts and Other New Techniques for the Administration of Libraries and Information Centers," *Library Resources and Technical Services*, 12:47–39 (Winter, 1968). Articles on some specific applications of the flowchart in libraries still regularly appear in the literature. A recent example is Hans H. Wellisch, "The ALA Filing Rules: Flowcharts Illustrating Their Application, with a Critique and Suggestions for Improvement," *Journal of the American Society for Information Science*, 34:313–30 (Sept., 1983).
15. Edward Yourdon, *Managing the Structured Techniques*, 2d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1979), p.149.
16. W. P. Stevens, G. J. Myers, and L. L. Constantine, "Structured Design," *IBM Systems Journal*, 2:115–39 (May, 1974).
17. Tom DeMarco, *Structured Analysis and System Specification* (Englewood Cliffs, N.J.: Prentice-Hall, 1979).
18. Chris Gane and Trish Sarson, *Structured Systems Analysis: Tools and Techniques* (Englewood Cliffs, N.J.: Prentice-Hall, 1979).
19. Gane and Sarson, *Structured Systems Analysis*, p.26.
20. Seymour Papert, *Mindstorms*, (New York, N.Y.: Basic Books, 1980). ■■

A Checklist of Issues to Be Considered Regarding the Addition of Microcomputer Data Disks to Academic Libraries

Diane Strauss

In recent years, library literature has reflected a growing interest in and enthusiasm for the new media that have recently become available to electronically store information. Magnetic computer tape for use with mainframe computers is being supplemented by magnetic and optical disk technologies for use with microcomputers, formats such as floppy and hard disks, laser videodiscs, and compact optical disks. As use of these formats becomes increasingly widespread, there has been a corresponding increase in the number of reference and research materials available on disk. Some duplicate printed sources, others are subsets of data files on computer tape, and still others are available only in disk format; titles range from the U.S. Census Bureau's *County and City Data Book* to the National Data Service's *College & University Library Survey*. As publication of these so-called data disks becomes commonplace, and as faculty and students become increasingly sophisticated in the use of microcomputers, many academic libraries will need to decide whether to add data disks to their collections.

In July 1985, a task force was appointed to investigate existing machine-readable information storage technologies and to make recommendations regarding their incorporation into the Academic Affairs Library of the University of North Carolina at Chapel Hill. The task force began its work by conducting a literature search and found that, while several articles had been published about microcomputer technology and about public and staff use of microcomputers for word processing, database management, and spreadsheet functions, almost nothing had been written about the practical aspects of collecting and making data disks available to the public. During the meetings that followed, the task force identified and resolved several issues concerning the selection, housing, and accessibility of data disks. Ultimately it recommended that a separate machine-readable data file (MRDF) unit be created to house data disks and other machine-readable information collected by the library, that it be staffed by reference personnel, and that the data disks be available for use only within the library. Other libraries may choose to handle data disks differently, but whatever the final outcome, certain common issues concerning selection, acquisition and cataloging, hardware requirements, staffing and levels of service, circulation, and compliance with copyright and licensing restrictions must be resolved. Some of the most important ones follow.

SELECTION

1. Does the library's existing collection development policy adequately provide for the addition of machine-readable data files

Diane Strauss is head, Business Administration Social Sciences Reference Department, Davis Library, University of North Carolina at Chapel Hill.

to the library, or must a new section (or policy) be written to cover material in this format?

2. Does the existing (or new) policy address the following issues?
   a. Does it specify that data disks be compatible with existing hardware?
   b. Does it require that program software needed to access specific data disks be held by the library, or does it allow for such programs to be ordered with the data disks as part of the information package?
   c. Who should select data files? Bibliographers? Reference librarians? Faculty?
   d. Should money for the purchase of data disks come out of the regular materials budget, or should a separate budget line be created for their purchase?
   e. How should gift software be handled? Should it be accepted if it is not compatible with existing hardware or if it is not accompanied by documentation or users' guides?

3. If a separate unit within the library is created to handle data disks and other machine-readable data files, should that unit also assume responsibility for selecting and housing supplementary printed directories, guides, and indexes? What about popular microcomputing periodicals?

HARDWARE AND SUPPLIES

1. Does the library presently have public access microcomputers?

2. If not, does it plan to use microcomputers that are currently assigned to internal operations? How will it balance public use with internal use?

3. Can the library supply microcomputers dedicated to the use of data disks and other machine-readable data files? If new hardware is being ordered, can the microcomputers being considered for use with magnetic disks also be modified for use with other types of machine-readable data formats such as laser videodiscs and compact optical disks?

4. How much memory is required for the microcomputer? Are any special enhancements, such as graphics boards and math coprocessors, needed?

5. Has the library budgeted for supplies and maintenance as well as for the hardware itself?

ACQUISITION AND CATALOGING

1. Advertisements for data disks are often vague and occasionally misleading. Considerable preliminary investigation may be necessary before placing an order to determine that the library has (or can get) the necessary hardware and program software to run the data disk. Who should be responsible for making these determinations? The selecting unit? The acquisitions department? The systems office?

2. Should orders for data disks be handled differently from orders for other materials? Are there parallels with other nonprint items? What information should be included on the order card? (Possibilities include specific format, number of disks, the existence of accompanying print documentation, and whether the material being ordered is in the public domain).

3. How should duplicate requests for noncopyrighted data disks that allow for unlimited copying—for example, those published by the U.S. Census Bureau—be handled? Which departments or individuals within the library should be responsible for making duplicate copies for data disks for other library units?

4. What should the acquisitions department's level of involvement be in inspecting orders? Should it be limited to the printed invoice that accompanies the order, or should it extend to examination of the disks themselves? If the former, who should be responsible for examining the disks?

5. At what point (and by whom) should backup copies be made? Should they be made upon receipt of the order, after inspection of both invoice and disks, or only after the disks and the accompanying documentation have been cataloged and sent to the library unit that ordered them?

6. What special skills, if any, are necessary for the person involved in ordering and inspecting data disks? If special skills are required, should one person in the department be designated to handle all steps of the acquisitions process?

7. What level of cataloging and classification should be provided?

LOCATION

1. Should all data disks and machine-readable data files be housed in a single unit, regardless of subject content? What are the advantages and disadvantages of such an arrangement? What about branch libraries?

2. Can the library or library department planning to acquire and house data disks meet minimum requirements for space, wiring, lighting, ventilation, noise control, and security? Does the proposed site have room for expansion?

3. Should the MRDF unit(s) be adjacent to or part of a public services department such as circulation or reference? If so, can circulation of software and accompanying printed documentation be handled from the service desk?

4. Where should backup disks be housed?

5. How should scheduling of public use of microcomputers and software be handled? Should it be on a walk-in or an appointment basis? Should any one group of library users have priority over any other group?

6. What hours should the MRDF unit be open? If open during all of the hours that the library is open, should full service be provided at all times?

STAFFING

1. Who should be responsible for supervising the operation of a machine-readable data collection and/or unit? Do they have the necessary skills? For example,

   a. Do staff have the requisite public service skills? Are they adept at instructing people in the use of research materials? Are they familiar with related print materials? Do they understand research methodology?
   b. Do staff have the necessary technical skills? Do they understand microcomputer operations? Do they know standard microcomputer programs such as dBase II and Lotus 1-2-3 well enough to use them in conjunction with data disks? Can they work with ASCII files and perform file conversions if necessary?
   c. How skilled are staff in basic statistical concepts and operations? Will they be able to help users reformat and manipulate the raw statistical data contained in many disks?

2. If staff lack the necessary skills, are training opportunities available on campus? Is the library able to subsidize training of staff?

3. How many people are necessary to staff such an area? Should they be professionals or support staff? How much of the work can be done by student assistants?

4. If the unit is adjacent to or part of a public service department, how much will staff not directly affiliated with the MRDF unit need to know? How much assistance should they give?

LEVEL OF SERVICE

1. What level of service should be provided? Should it be limited to simply making hardware and data disks available, or should it extend to fairly sophisticated advice concerning statistical methodology and training in the use of programs needed to operate data disks?

2. What should the library's policy be regarding actually doing the research for library users, particularly those who lack basic microcomputing skills?

3. Who should be responsible for copying public domain data disks for library users' personal software libraries? Who should supply the blank disks for such copying?

4. To what extent should the library prepare supplementary documentation when printed documentation accompanying a data disk is vague or incomplete?

CIRCULATION

1. Should data disks (and accompanying documentation and program disks) be allowed to circulate outside the library?

2. If so, how can the library avoid damage to disks by careless users and/or library staff?

3. To what extent should returned data disks be inspected? How will staff discover if they have been damaged or if parts of data files have been erased?

4. Should the library set up a fee schedule to recover the cost of lost or damaged data disks?

5. Has the library's circulation policy been amended to include data disks and other machine-readable materials?

COPYRIGHT AND LICENSING

1. Who should be responsible for determining any copyright and/or licensing restrictions prior to ordering? Acquisitions? The selecting unit? The systems office?
2. Should the library take a passive role (for example, posting signs) or an active role (for example, monitoring patron use) in enforcing copyright and licensing restrictions? How best can the library comply with the law?
3. Should public domain disks be made visibly distinguishable from those with copyright and licensing restrictions?

OTHER

1. What other campus organizations are collecting data disks or machine-readable data files? Does a referral network need to be established?
2. How can the library and other campus organizations work together regarding training in basic microcomputer literacy and/or use of specific programs? Can cooperative collection development policies be established?
3. How should the library publicize the addition of data disks and accompanying services to potential users?

Adding data disks and other machine-readable data files to a library collection requires a substantial commitment in library staff, resources, and time, but with answers to some of the questions that have been mentioned above, libraries can be well on their way to deciding whether to add such materials to their collections and how best to make them available to library users.

Note: The final report of the University of North Carolina at Chapel Hill's Task Force on Machine-Readable Data Files has been submitted to the ERIC Clearinghouse on Information Resources. ■■

Communications

LCS Shelflist Maintenance Using an OCLC M300

John Drueke

Lovejoy Library, along with twenty-six other academic libraries in Illinois, uses the Library Computer System (LCS), an online bibliographic network administered by the University of Illinois, for circulation, interlibrary loan, and shelflist searching by library users. The LCS database stores the Library of Congress call number, author, title, place and date of publication, Library of Congress card number, circulation status, and holdings records for 441,000 of Lovejoy's titles. A system user can search the shelflist for an item by author, title, author-title, or call number; in addition, a copy of a holding can be charged, renewed, saved, snagged, or discharged. Because LCS' most convenient method of performing these circulation transactions uses an item's call number and copy number to form a unique identifier, the circulation desk employee must know the *exact* call number as it is stored in the machine-readable shelflist to successfully charge or discharge a book. The usual source for the call number is the book's spine label, which is presumed to show that identifier. When a circulation transaction is attempted with a nonexistent call number, LCS responds with "CALL NUMBER NOT IN LIBRARY SHELFLIST," and the transaction is cancelled.

DECIMAL POINT PROBLEM

Unfortunately, when the library's shelflist was converted from 3-by-5-inch cards to LCS machine-readable form by a commercial service in 1981, the decimal point before the Cutter number was omitted on approximately 80 percent of the call numbers. Records added to the LCS shelflist since then do contain the required decimal point. This inconsistency in call number format slows circulation transactions, as well as other call number searching, because the terminal operator cannot know beforehand whether the LCS record for the item at hand contains the decimal point. The LCS call number error message cannot be accepted as final until both forms of the call number are entered. Many circulation transactions, therefore, must be initiated twice, increasing service time and the frustration of both staff and borrowers.

SOLUTIONS

A custom program designed to directly access each record in our shelflist and automatically insert the decimal point where required is a low priority for LCS because the problem affects only Lovejoy Library and not the entire system. The only method for call number correction available to the library staff has been to search systematically the online shelflist noting those call numbers lacking the required decimal point. Corrections to those call numbers must then be made at a terminal, one at a time, using Superwylbur, a line-oriented text editor that runs on the same mainframe computer as LCS. Change requests entered through Superwylbur are not made directly to the shelflist but are stored in a file at the mainframe and executed by a weekly batch maintenance program. Because this editing process is so time-consuming, library student staff has corrected only classifications *A* to *BL* in the last two years. To increase the effectiveness of this correction procedure, we decided to automate it using an available OCLC M300 terminal operating as a stand-alone microcomputer. Our

John Drueke is systems coordinator, Lovejoy Library, Southern Illinois University at Edwardsville.

goal was to search our entire online LCS shelflist, find the improperly formatted call numbers, and transmit corrections to Superwylbur, all with the least possible human intervention.

HARDWARE

The OCLC M300 terminal, a modified IBM PC, has 256 kilobytes of memory, two 5¼ inch diskette drives, and an asynchronous communications interface (PORT 1 on the rear of the computer). The interface was connected to a 1,200 bit per second (bps) LCS data line previously used with a standard LCS terminal. No modifications to either the M300 or the LCS line were required.

SOFTWARE

Since no appropriate software was available, two programs were written in-house using IBM PC BASIC, version A2.10. BASIC was well suited to our purpose because it provides communication facilities for use with the asynchronous adapter on the M300, character string manipulation routines necessary for easily parsing and constructing call numbers, and quick debugging and modification of small programs. In addition, it is supplied at no extra charge with every M300, and a library staff member had BASIC programming experience.

The first program, called the Proofreader, was designed to find incorrectly formatted call numbers. It requests a range of LCS title (accession) numbers from the operator, logs on to the system, and sends the appropriate LCS command to retrieve each bibliographic record in the specified range. As each response is received by the M300, the Proofreader parses the call number and determines that it falls into one of three categories:

- A correctly formatted Library of Congress call number with the decimal point in place. These are counted for statistical purposes but otherwise ignored.
- An LC call number missing the decimal point but otherwise correct. These are counted and stored in a diskette file for later automated processing.
- A series of characters that does not comprise a valid LC call number, for example, 9876H5, 123.56, HVD47. These are counted and recorded in a second diskette file for review by a staff member and possible manual correction.

The Proofreader can run unattended after office hours and can search and process approximately 450 LCS bibliographic records per hour. The limiting factor in the proofreading rate is the 1,200 bps data line. Successive runs of the Proofreader are executed until several thousand improperly formatted call numbers are collected in the diskette file.

The second program, the Decimal Point Editor, is then run; it logs on to Superwylbur, the system text editor, and sends the required commands to begin a maintenance session. The program then transmits each of the incorrect call numbers stored in the disk file, a newly constructed call number containing the decimal point, and the appropriate editing commands. When all call numbers in the diskette file have been processed, the program logs off Superwylbur. The Decimal Point Editor can send approximately 2,150 call number corrections to Superwylbur per hour. At the scheduled time, the weekly LCS maintenance program uses the data transmitted to Superwylbur to make the actual changes in the online shelflist.

EVALUATION

LCS has limited Lovejoy to twenty thousand call number corrections per week because processing time and mainframe memory requirements for such large numbers of corrections can be excessive. At that rate, all call numbers in the shelflist will be proofread and corrected, if necessary, in approximately twenty-two weeks. While this method requires considerably more time than direct manipulation of the database, it makes few demands on system resources and library staff time. Furthermore, it has been error free in automatically constructing and transmitting the correct call numbers. An estimated $20,000 in student wages may be saved by automating the correction process. In addition, we now have a methodology for using the microcomputer to supplement or re-

place the maintenance tools provided by the online system. Similar processes can be used for other large-scale operations that the library or the system would otherwise be unable to undertake.

BIBLIOGRAPHY

Hart, Elaine, ed. *LCS Reference Manual.* (Chicago:1982).

International Business Machines Corporation. *BASIC Reference.* 3d ed. IBM Personal Computer Hardware Reference Library. (Boca Raton, Florida:1984).

OCLC Online Computer Library Center, Inc. *OCLC M300 Guide to Operations.* 2d ed. Microcomputer Series. (Dublin, Ohio:1985). ■■

Using a Serials Control System in a Reference Setting

Thomas E. Smith

INTRODUCTION

This paper will focus on how PHILSOM is used as a serials subsystem for Reference services at the Paul Himmelfarb Health Sciences Library at George Washington University Medical Center. What is remarkable is that PHILSOM, as a serials control system, is not designed for direct public services use. PHILSOM is user friendly to serials librarians who use it daily but not necessarily user friendly to those who use the system infrequently. At GWUMC, however, PHILSOM has been used by Reference librarians to help library users in their quest for information about serials and serial holdings. The experience at GWUMC suggests that other available serials control systems now used mainly in a technical services setting could be used by Reference departments elsewhere to provide an interactive environment for Reference librarians.

Thomas E. Smith is serials librarian, Paul Himmelfarb Health Sciences Library, George Washington University Medical Center, Washington, D.C.

PHILSOM

Since 1962 PHILSOM has established itself as one of the premier medical serial control systems and biomedical bibliographic databases in the United States. The database, developed at the Washington University School of Medicine Library in St. Louis, was operating in a network environment by 1968.[1] PHILSOM is an acronym for Periodical Holdings In Libraries of Schools Of Medicine.[2] The online programs for PHILSOM are written in standard MUMPS. Now in the fourth generation of software, the PHILSOM IV database contains records for more than sixteen thousand biomedical titles. Member libraries may access the database in one of three ways:

1. Dialing into PHILSOM headquarters in St. Louis via dedicated lines or Telenet. The WUSML computer can handle access by multiple libraries. The data for each library is stored in separate files. No library has *write* access to another library's data. *Read* access is available to all network libraries and the PHILSOM office.

2. Leasing the software and using the IBM PC/AT microcomputer or equivalent.

3. Leasing the software and using one's own minicomputer. At GWUMC we run the PHILSOM software on a DEC PDP 11/44.

The last two options are referred to as *distributed* systems.

The PHILSOM Record

Each PHILSOM record has two parts: a bibliographic file (Basic Data) and location data specific to each title (Local Data). A third file, Accounting Data, can be accessed as well. Basic Data is controlled by St. Louis; Local Data by individual network libraries. The following are fields for each file and their tags. Please note the PHILSOM record does not follow the MARC-S format (See figures 1 and 2).

Basic Data

System control number (ALPHA NUMBER)
Short title (SHORT)
Full title (FULL)
Cataloging level (CAT LEVEL)

```
------------------------------ BASIC DATA ------------------------------
ALPHA NUMBER:002420003          SHORT:ACTA HISTOCHEM Y1:1954 ISSN:0065-1281
FULL1:ACTA HISTOCHEMICA
CAT LEVEL:1 STATUS:1 V:82 I:2 IV:2 FREQ CLASS:BM SB1:256 LAN1:GER IND1:1
IND2:4 IND3:3 IND4:5 IND5:10 SERLINE:A05865000
FREQ: JAN=0 FEB=1 MAR=0 APR=1 MAY=0 JUN=1 JUL=0 AUG=1 SEP=0 OCT=1 NOV=0 DEC=1
XREF1:300031003 ACTA HISTOCHEM (JENA)
SET# 1------------------------ LOCAL DATA --------------------GWUO:15-JUL-85
LAST UPDATED BY GWU ON 19-JUN-85 AT  3:06 PM
ACT:4 VOLUM:75 ISSUE:2 DATE:30-JAN-85 IL:02 BC:X COL:N STR:2 COV:N TP:L IN:Z
ADS:+ FR2:R SOURCE:80 EXP DATE:123185 BNDR:H GAP:C LATE:70
HOLD1:1 1963 1984/15-75/
COML1:VOLS. 15-75  BOUND AS OF 19-JUN-85
COML2:SHIPPED: VOL. 74-75, 1984; AVAIL. 20-JUN-85
NOTE1:**CLAIM: V76 N1 (5-MAY-85)
```

Fig. 1. Example of a PHILSOM Record from the Work Copy. This is an Active Title.

```
------------------------------ BASIC DATA ------------------------------
ALPHA NUMBER:002310003          Y1:1948 Y2:1968
FULL1:ACTA GENETICA ET STATISTICA- MEDICA
BIB 1:CONT AS HUMAN HEREDITY WITH V. 19, 1969.
CAT LEVEL:1 STATUS:2 SB1:222 SB2:701 SERLINE:A05580000
SET# 1------------------------ LOCAL DATA --------------------GWUO:15-JUL-85
LAST UPDATED BY JIM ON 2-APR-84 AT  2:21 PM
ACT:3 IL:02 FR2:R
HOLD1:2 1960 1968/10-18/
COML1:LIBRARY HAS SUPPLS. TO VOLS. 10,15-17
COML2:VOLS. 10-18 ARE BOUND
```

Fig. 2. Example of an Inactive Record.

Activity Status (STATUS)
Volume (V)
Issue (I)
Issues per volume (IV)
Frequency (FREQ CLASS)
Subject Codes (SB)
Language codes (LAN)
Where Indexed (IND)
Serline number (SERLINE)
Publication schedule (FREQ)
Cross References (XREF)

Local Data

Date record last updated
Active or Inactive Record (ACT)
Last volume checked in (VOLUM)
Last issue checked in (ISSUE)
Date of Check in (DATE)
Internal Shelving Location (IL)
Bindery Information (several tags)
Routing Information (RT, ROUT)
Vendor source (SOURCE)
Expiration Date (EXP DATE)
Late Field (LATE)
Note field (NOTE)
Comment fields (COML) These print on the paper copies

The PHILSOM records can be accessed by the ISSN number, alpha number, title key, or a keyword search.

User Aids

Although PHILSOM is essentially a serials control system, network libraries have a number of products designed for public use. In particular, there is the serials holding list or *desk copy*, which has only the barest bibliographic information (from the Basic Data) and holdings information plus comment lines (from the Local Data). It is updated monthly. (See figure 3). Another user aid is the *KWOC list*, (Key Word Out of Context) which is produced annually by PHILSOM headquarters for each library, based on the participating library's title holdings. (See figure 4). A subject list can be produced locally by member libraries any time. (See figure 5). A list by language also is available.

FORTSCHRITTE DER KARDIOLOGIE
INCLUDED IN BIBLIOTHECA CARDIOLOGICA, V. 6, V. 9, V. 12
AS FORTSCHRITTE DER KARDIOLOGIE-ADVANCES IN CARDIOLOGY-
PROGRES EN CARDIOLOGIE. SUSP 1963 - 1969. CONT AS
ADVANCES IN CARDIOLOGY WITH V. 4, 1970.

GWUO: 1959 1962: 2-3
LABELLED AND SHELVED AS ADVANCES IN CARDIOLOGY
ALL HOLDINGS ARE BOUND

FORTSCHRITTE DER KINDERCHIRURGIE
SEE:
PROGRESS IN PEDIATRIC SURGERY

FOUNDATION DIRECTORY

GWUO: 1983 1983: 9
SHELVED IN REFERENCE AS 911 .A2F65
LIBRARY RETAINS CURRENT EDITION ONLY

FRONTIERS OF BIOLOGY

GWUO: 1966 1979: 1-3/5-10/12/14-16/19/22-24/26-27/29-34/36-39
41-48
CONSULT CARD CATALOGUE FOR INDIVIDUAL TITLES AND
CALL NUMBERS.
SENT P&A TO BACK AUG 30,1984 (VOL 40)

FRONTIERS OF HORMONE RESEARCH

GWUO: 1972 1984: 1-12
CIRCULATES TWO WEEKS
ALL HOLDINGS ARE BOUND

FRONTIERS IN NEUROENDOCRINOLOGY
SUSP 1974 - 1975.

GWUO: 1969 1973: 1-3
ALL HOLDINGS ARE BOUND. STANDING ORDER DISCONTINUED.

GANN
CONTINUED AS JAPANESE JOURNAL OF CANCER RESEARCH
WITH V.76, 1985.

GWUO: 1974 1984: 65-75
VOLS. 65-74 BOUND AS OF 2-AUG-84

GANN. JAPANESE JOURNAL OF CANCER RESEARCH
SEE:
GANN

Fig. 3. Sample from the PHILSOM Desk Copy.

The *work copy* frequently is used as a backup in the Technical Services area. It contains complete records for all titles held by the library (both Basic and Local Data). Older copies are kept at the reference desk in case of down time. The *Union List Microfiche* is available for both Serials and Public Services staff to determine the holdings information for titles held by other PHILSOM members.

PHILSOM at GWUMC

The Himmelfarb Library has been a distributed library since 1980. Very early after the implementation of PHILSOM as our serials' subsystem, we began to explore possible uses for PHILSOM outside the Serials Department. One hurdle to overcome was the sign-on procedure for PHILSOM functions, (MSL:M). It consisted of a location

```
OPINION
   CURRENT MEDICAL RESEARCH AND OPINION
   PSYCHIATRIC OPINION


OPPORTUNITIES
   MINORITY STUDENT OPPORTUNITIES IN UNITED STATES MEDICAL
     SCHOOLS


OPPORTUNITY
   OPPORTUNITY PLACEMENT REGISTER


ORAL
   INTERNATIONAL JOURNAL OF ORAL SURGERY
   JOURNAL OF ORAL AND MAXILLOFACIAL SURGERY
   JOURNAL OF ORAL MEDICINE
   JOURNAL OF ORAL SURGERY
   ORAL SURGERY, ORAL MEDICINE, ORAL PATHOLOGY
   POPULATION REPORTS. SERIES A. ORAL CONTRACEPTIVES


ORGANIC
   JOURNAL OF THE CHEMICAL SOCIETY.  PERKIN TRANSACTIONS.
     1.  ORGANIC AND BIO-ORGANIC CHEMISTRY
   JOURNAL OF THE CHEMICAL SOCIETY.  PERKIN TRANSACTIONS.
     2.  PHYSICAL ORGANIC CHEMISTRY
```

Fig. 4. Sample Listing from the KWOC List.

designator, name and password, and, after a brief fortune (as a reward for signing on correctly), the user would be prompted to declare what function was desired. There are, however, almost fifty functions to choose from, each user having only certain assigned functions. Since persons outside the Serials Department would only need to look at the records for both titles we owned and did not own, this sign-on procedure was cumbersome at best. The sign-on procedure, therefore, was abbreviated. This new log on displayed only serials data necessary for librarians working in a public services capacity. Function "PD" (PHILSOM DISPLAY) was an attempt to make PHILSOM "librarian friendly." PHILSOM could now be accessed online at the reference and circulation desks to give both users and librarians updated and current information about bindery shipments, incomplete journals, holdings data, and the date the last issue of a journal was checked in, to name a few. Later, access to circulation data was added. Since journals five years old and older may circulate, this information is quite valuable. Function PD also reduced the number of questions referred to the Serials Department. Patrons benefitted by having unambiguous responses to their inquiries. Public Services librarians also had up-to-the-minute information about any title held by the library.

The Reference Link

Using PHILSOM has unquestionably enhanced the quality of reference service we provide our users. Our monthly reference statistics repeatedly show that questions concerning periodicals and holdings constitute the single largest category of questions asked. Without online access to PHILSOM providing users with up-to-date serials, information would be difficult at best.

Why then use what is essentially a serials department database in a reference setting? The advantage of an interactive system is the type and currency of information avail-

```
HIMMELFARB LIBRARY SUBJECT LIST

ANATOMY

001870003: ACTA ANATOMICA

001880003: ACTA ANATOMICA. SUPPLEMENT

007410003: AMERICAN JOURNAL OF ANATOMY

009160003: ANATOMICAL RECORD

018480003: BIBLIOTHECA ANATOMICA

019310003$ BIOLOGICAL STRUCTURE AND FUNCTION

019860003$ BIRTH DEFECTS ORIGINAL ARTICLE SERIES

020630003$ BRAIN, BEHAVIOR AND EVOLUTION

032460003: CURRENT TOPICS IN MEMBRANES AND TRANSPORT

044690003$ IN VITRO

047480003: INTERNATIONAL REVIEW OF CONNECTIVE TISSUE RESEARCH

050400003: JOURNAL OF ANATOMY

052630003$ JOURNAL OF EMBRYOLOGY AND EXPERIMENTAL MORPHOLOGY

054860003: JOURNAL OF MORPHOLOGY

092750003$ TECHNIQUES OF BIOCHEMICAL AND BIOPHYSICAL MORPHOLOGY

093470003$ TISSUE AND CELL

098392003$ VIRCHOWS ARCHIV. A. PATHOLOGICAL ANATOMY AND
           HISTOPATHOLOGY
```

Fig. 5. Listing from the Locally Produced Subject List.

able. Reference librarians are able to view the comment lines (COML) that are intended for public view. Information about bound holdings, bindery shipments, and whether a title is in the reserve collection can be found in comment lines. However, information not necessarily intended for public viewing—information that does not print out in the paper products—can be found in the online note fields (NOTE). Librarians also find this data very helpful. Information on erratic publication patterns, recent claiming history, incomplete volumes, and updates from publishers and vendors can be gleaned from viewing these notes for internal consumption. Since we have a read-only system, there is no security problem since changes to the location data can only be made in the serials department.

As with any system there are some areas

of controversy. First and foremost, PHILSOM is designed to be a serials control system. When Function PD was introduced to the staff, no modifications were made in any part of the record. The basic data, which is not in the customary "card image," is initially difficult to read. The PHILSOM field designators are mnemonic and often are confusing to non-serials staff who don't use the records every day.

Another source of confusion is subject headings. PHILSOM formerly used UCMP subject headings. Many libraries found them obsolete and they were switched to subject headings found in NLM's *List of Titles Indexed in Index Medicus*. MeSH headings were added as a separate category over and above the NLM headings. Many PHILSOM members found MeSH headings too specific for collection-development purposes, and so they were dropped in 1985. The NLM headings are the only subject headings that remain. In Function PD these headings are indicated by a three-digit number. This three-digit number in turn must be looked up in another manual. The only way to get around this procedure is to either consult the paper subject list or sign off the local, librarian-friendly access and go into the unaltered programs using the Function "SJ" (Subject Search). With this function added to the main PHILSOM menu in 1984–85, a reference librarian can get keyword access to journals. Function "SJ" was designed for patron and reference access and has four levels of detail available.

Another concern is the nonstandard format. Some librarians would prefer to see adherence to national standards (MARC, MeSH) as opposed to system-specific standards. It also has been noted that sometimes there is too little bibliographic information in the basic data. Some librarians would like to see a fuller record. There is also the danger of presenting too much information as well as giving too little information.

CONCLUSION

This paper has shown how a serials control system can successfully be used for reference service. The advantages to having online access to PHILSOM were instantly obvious: elimination of manual records for binding shipments and incomplete journals, latest issue received of a journal, and up-to-date holdings information. The disadvantages were confusion about the tags and subject headings. Adherance to national formats can be a reassuring feature to many librarians who have worked with OCLC, LC, or NLM products.

Our experience using PHILSOM at GWUMC and making it available to staff online is generally favorable. Other libraries, using different serials control systems (LINX, PERLINE, REMO, INNOVACQ) exclusively in a Technical Services setting may wish to consider extending their system applications to a Reference/Public Services environment.

REFERENCES

1. Judy McQueen and Richard Boss, "Serials Control in Libraries: Automated Options," *Library Technology Reports* 20, no. 2:258–59 (Mar./Apr. 1984).
2. Susan Crawford and others, "Technology at Washington University School of Medicine Library: BACS, PHILSOM and OCTANET," *Bulletin of the Medical Library Association* 71, no. 3:324 (July 1983). ■■

Current Status and Future Trends of East Asian Library Automation in North America

Karen T. Wei

The implementation of the Research Libraries Group's (RLG) Chinese, Japanese, and Korean (CJK) automated system for the processing of East Asian language materials has been one of the most welcome achievements in modern library technology in the area of Asian librarianship. On Sep-

Karen T. Wei is assistant Chinese librarian, University Library, University of Illinois at Urbana-Champaign.

tember 12, 1983, the Library of Congress entered their first online cataloging record containing Chinese vernacular script into the RLIN (Research Libraries Information Network) database.[1] Many major East Asian libraries in North America joined in this venture, which inaugurated a new era of online resource sharing. This breakthrough finally allowed East Asian libraries to input, store, maintain, retrieve, and output their CJK materials in vernacular languages and scripts in library computers in the same way Western-language peers had done for years.

Although most large East Asian libraries are currently taking part in the RLG CJK project, many medium-sized and small libraries have chosen to wait for the results of this development before deciding to participate in the system. In addition, the development of the OCLC (Online Computer Library Center) CJK system may contribute to the lack of participation of small libraries. The present survey was designed to gather data on the current status of automation in East Asian libraries and to assess the projected trends in this field. In November 1984 a three-page questionnaire was sent to the eighty member libraries listed in the *Directory of East Asian Collections in North American Libraries*[2] and to the Los Angeles County Public Library Project Asia (See appendix). A brief cover letter accompanied each questionnaire. Followup letters were sent out in January 1985 and again in March 1985 to ensure a greater number of responses. With seventy libraries (or 87 percent) responding to the survey it is possible to state that the current status of East Asian library automation is as follows.

SURVEY SUMMARY

The respondents can be divided into two groups: (1) current RLIN CJK users (eighteen at the time of the survey) and (2) non-CJK users (sixty-three). Among the eighteen CJK users, seventeen (or 94 percent) libraries returned the questionnaire. The majority are college and research libraries, with the exception of the Library of Congress and the Los Angeles County Public Library Project Asia (a library acquisition/cataloging service center). Leaving aside the Library of Congress, twelve of the users are RLIN full members (i.e. they use the RLIN automated system for bibliographic control of their main collections) and the remaining four use RLIN only for CJK collections, thus, constituting an "associate member" group that uses other forms of automated system for their main collections. Most of the libraries have separate East Asian collections and some reported having separate East Asian reference collections; others are administratively separate from, but physically integrated with, their main collections. Two libraries have CJK collections of more than one million volumes, eleven have between one hundred thousand and one million volumes, and four have less than one hundred thousand volumes. Depending on the size of the library, the participants employ one to four CJK terminals (except the Library of Congress, which has twenty-six terminals) for the processing of Chinese, Japanese, and Korean language materials. Two participants use both the RLIN acquisitions and cataloging system for the processing of CJK materials; all the other libraries currently use the system for cataloging and limited reference services only.

The range of starting dates for inputting was September 1983 to February 1985. With a combined total of sixty-six terminals in use to date, the database has grown to more than one hundred thousand records in the books file and more than two thousand records in the serials file (the two most heavily used files) since September 12, 1983.

During the first year-and-a-half of the CJK operation, four libraries indicated that their cataloging statistics were improving and one reported a 100 percent increase in production. Three libraries stated that their cataloging activity remained the same as before joining the system, seven of the libraries have experienced a decrease in their cataloging activities for a variety of reasons, and two libraries are too new to the system to know the effect as yet. Although the Library of Congress has been doing the lion's share of original cataloging added to the RLIN database (especially after De-

cember 1984 when LC went full force in online cataloging), other libraries have tried to contribute as many original cataloging records as possible. The percentage of original cataloging by the non-LC libraries varies from 10 to 95 percent (per institutional reports) while the derived records vary from 5 to 90 percent. These figures, of course, are highly dependent upon the library's policy.

There were numerous concerns about the card catalog before the CJK was in full operation. Questions of whether the card catalog should be closed were raised. If so, when? If not, should the current card production remain as it is or should other measures be taken into consideration? If the library were to subscribe to the RLG CJK printed cards, what should be done about the East Asian characters that are missing from the cards? How should the RLG cards be filed? That is, should they be interfiled in the existing card catalog or in a separate card catalog? Among the seventeen libraries that responded to the survey, four are completely online, three of which no longer file cards. The fourth subscribes to RLG shelflist cards only. Eleven libraries continue to file LC and other locally produced catalog cards, and two subscribe to RLG CJK printed cards. There are no East Asian characters on the latter cards; one library writes CJK scripts on main entry cards only while the other library does not write the characters at all. Both libraries interfile the RLG cards into the existing card catalogs. While one library recognizes the quality and usefulness of the RLG cards, the other library indicates that the CJK characters would be more helpful.

Among the fourteen libraries that do not subscribe to RLG CJK printed cards, six strongly indicated that they do not intend to buy the cards in the future (apparently to clear the way for full automation); one library will buy RLG cards only when the vernacular characters are available. The remaining seven libraries are either undecided at present or are waiting for the outcome of the joint LC and RLG pilot project, which is supposed to produce cards with both CJK scripts and romanized forms. It is interesting to note that, despite the moderate cost of the RLG CJK printed cards with the convenience of complete headings on each set, the majority of the libraries still continue to use either LC cards or to produce their own cards. The survey also confirms the fact that, if the cards are to be continued, CJK scripts on the printed cards are extremely important to the users of the East Asian libraries.

Approximately half (eight) of the RLG libraries indicate that the public has access to the CJK terminals, mainly through the reference services of the library while the others (nine) say the public has absolutely no access to the terminals. It is important to note that the patrons of those four libraries that are completely online have access to the CJK terminals. In addition, four libraries surveyed indicated that they will eventually convert their older catalog files into the RLIN database, the rest will not or will consider it only if the money becomes available. Public access to the CJK terminals and retrospective conversion of data are the two major factors if a library intends to be completely online. Otherwise the public will have two files to search—the old card catalog and the new online catalog.

Fifteen of the seventeen CJK respondents stated that they are satisfied with the RLIN CJK system while the other two libraries indicate the system is only fair. About half (nine) of the libraries believe the cost to maintain the system is too expensive overall, while the other half (eight) thinks the cost is fair, taking into consideration all the factors involved. However, despite the opinion about the cost, two-thirds (twelve) of the current RLIN CJK participants intend to stay with the system even if the OCLC CJK system comes up. Others take a wait-and-see attitude or do not know at present what they will do when the OCLC CJK system comes up. However, the libraries that intend to stay with the RLIN CJK are largely RLIN full members. Therefore, the chances that these libraries will switch to OCLC is rather remote unless their main collections also change to that system. One other important factor is that the current RLIN CJK participants constitute approximately 58 percent of the

CJK collections nationwide.[3] It would appear to be wise and beneficial to stay where the major collections and records are.

The second group of survey respondents are non-RLIN CJK users, a total of sixty-three, of which fifty-three (or 84 percent) returned the questionnaire. They are made up of thirty-five college and university libraries, two public libraries, and sixteen museums and special libraries. Among this group, thirty-five of the libraries currently use the OCLC automated system for bibliographic control; seven use the RLIN system; two use the UTLAS system; three others use local systems; and five reported that they did not employ any automated system at all.

Thirty-three of the libraries in this group have separate East Asian libraries while the other twenty have no separate collections. Eleven libraries have an East Asian collection of one hundred thousand to one million volumes; twenty-nine have ten thousand to one hundred thousand volumes, and twelve have fewer than ten thousand volumes. Although none of the libraries are currently RLG CJK members, eight libraries have indicated their interest in joining the system, some of which will do so within the next couple of years. The major reason for not joining at the present is largely the budgetary problem, i.e. lack of funds for operating the system. Besides the eight potential RLG libraries, ten of the remaining libraries will join the OCLC CJK system; twelve libraries have no interest in joining either one of the automated system for East Asian collections, and the remaining twenty-three libraries are undecided (see table 1). In the undecided group, eleven are not decided on either system, six indicated that they are not interested in the OCLC CJK system but are undecided about RLIN CJK system, three libraries will consider OCLC CJK but are undecided about RLIN CJK, and three definitely say no to the RLIN CJK system but are undecided on the OCLC CJK system (see table 2).

Table 1. Non-CJK Users Group (N = 53)

System	Number
will join RLIN CJK	8
will join OCLC CJK	10
will not join either	12
undecided	23

Table 2. Undecided Group (N = 23)

System	Number
undecided on either system	11
undecided on RLIN CJK/not interested in OCLC CJK	6
undecided on OCLC CJK/not interested in RLIN CJK	3
undecided on RLIN CJK/will consider OCLC CJK	3

CONCLUSION

Although the current RLIN CJK users amount to a total of twenty, the collections of these libraries comprise approximately 58 percent of the total CJK collections nationwide. The majority of the CJK libraries are satisfied with the RLIN CJK system and intend to stay with the system regardless of the outcome of the OCLC CJK development. A small number of libraries, however, have an opportunity to decide whether to stay with the system when OCLC CJK comes up. Leaving aside the twelve libraries that have no intention of joining either RLIN or OCLC system in the future, the potential RLIN CJK and OCLC CJK users are divided among the non-CJK participants.

The respondents pointed out that the cost of operating the RLIN CJK system is a major factor that would affect their decision as to which system they will eventually join. Although it is too early to predict the future of either system, the development of the OCLC CJK will certainly play a crucial role in the decision making of medium-sized and small East Asian libraries. If the quality of the OCLC CJK system is equal to or greater than the RLIN CJK but has a cheaper rate, the potential for the OCLC CJK system is good. Otherwise RLIN CJK system will continue to dominate the East Asian library community in technical processing and resource sharing for many years to come.

REFERENCES

1. *Library of Congress Information Bulletin* 42:347 (October 10, 1983).

2. *Directory of East Asian Collections in North American Libraries* (Washington, D.C.: Committee on East Asian Libraries, Association for Asian Studies, 1983).
3. Based on the final report of "Current Status of East Asian Collections in American Libraries 1979/80," *CEAL Bulletin* 70/71:68–81 (Feb/June 1983).

APPENDIX. SURVEY QUESTIONNAIRE

1. Please indicate the type of your library
   _____ College and university library
   _____ Public library
   _____ Federal library
   _____ Museums and special libraries
   _____ Others, please specify _____
2. Please indicate the automated system that your library uses for bibliographic control
   _____ OCLC
   _____ RLIN
   _____ WLN
   _____ UTLAS
   _____ Others, please specify _____
3. Do you have a separate East Asian collection?
   _____ Yes
   _____ No
4. What is the size of your CJK collections?
   _____ More than 1,000,000 volumes
   _____ 100,000 to 999,999 volumes
   _____ 10,000 to 99,999 volumes
   _____ Less than 10,000 volumes
5. What is the size of your staff (FTE)?
   Number of staff
   _____ Professional
   _____ Clerical and technical assistant
   _____ Others, please specify _____
6. Does your East Asian library currently participate in the RLG CJK automation project?
   _____ Yes
   _____ No
   If your answer is yes, please go to questions 7–16;
   If your answer is no, please go to questions 17–19.
7. If your library is currently participating in the RLG CJK automated system
   a. When did you join the system _____ (date)
   b. When did you first input your online acquisition/cataloging records into the RLIN database _____ (date)
   c. How many terminals do you have and what are their functions?

Number of terminals	Function
_____	Acquisition
_____	Cataloging
_____	Reference
_____	All the above
_____	Others, please specify _____

   d. What is your average monthly acquisition/cataloging input into the RLIN system?

Acquisition	Cataloging	
_____	_____	More than 3,000 titles
_____	_____	2,000 to 2,999 titles
_____	_____	1,000 to 1,999 titles
_____	_____	500 to 999 titles
_____	_____	Less than 500 titles

   e. What are the percentages of your cataloging record for
      _____ Original cataloging
      _____ Derived, upgraded
      _____ Derived, not upgraded
      _____ Others, please specify _____

f. What were the average monthly cataloging statistics before joining the system?
   _____ More than 3,000 titles
   _____ 2,000 to 2,999 titles
   _____ 1,000 to 1,999 titles
   _____ 500 to 999 titles
   _____ Less than 500 titles
g. The percentage of cataloging increase or decrease since joining the project
   _____ Percentage increase
   _____ Percentage decrease

8. If your library is currently participating in the RLG CJK automated system, what kind of printed cards do you use?
   _____ No cards used; completely online
   _____ Subscribe to RLG cards only
   _____ Continue to use LC cards
   _____ Subscribe to both LC and RLG cards
   _____ Produce own cards
   _____ Other(s), please specify _____
9. If you are subscribing to the RLG CJK catalog cards, what do you do with the CJK scripts that are missing from the cards?
   _____ Do not write CJK characters
   _____ Write characters on all cards
   _____ Write characters on selected cards, please specify _____
   _____ Others, please specify _____
10. How do you file the RLG cards?
   _____ File separately from the existing card catalog
   _____ Interfile with the existing card catalog
   _____ Others, please specify _____
11. If you do not currently subscribe to the RLG cards, do you plan to do so in the future?
   _____ Yes
   _____ No
   _____ Undecided
   _____ Others, please specify _____
12. Does the public have access to the CJK terminals in your library?
   _____ Yes
   _____ No
13. If you are currently cataloging CJK materials online into the RLIN database, do you plan to convert the older records into RLIN in the future?
   _____ Yes
   _____ No
   _____ Other(s), please specify _____
14. What is your opinion of the RLG CJK automated system?
   _____ Satisfactory
   _____ Fair
   _____ Not satisfactory
   _____ Other, please specify _____
15. What is your opinion of the cost of maintaining the system?
   _____ Too expensive
   _____ Fair
   _____ Not expensive
   _____ Other, please specify _____
16. If you are a current RLG member, what will you do when OCLC East Asian program becomes available?
   _____ Stay with RLG CJK program
   _____ Switch to OCLC East Asian program
   _____ Retain RLG and use the OCLC system
   _____ Other, please specify _____
17. If you are not a current member of the RLG CJK project, do you plan to become one in the future?
   _____ Yes
   _____ No
   _____ Undecided

18. If you plan to participate in the RLG CJK project, please indicate the possible date _____ (date) and give the reason for not joining now _____
_____

19. If you do not plan to participate in the RLG CJK program, will you consider joining the OCLC East Asian program when it becomes available?
_____ Yes
_____ No
_____ No intention to participate in any form of automated system for East Asian collection
_____ Other(s), please specify ■■

LIBRARY & INFORMATION TECHNOLOGY ASSOCIATION

. . . the only professional association devoted exclusively to the use of technology in libraries.

APPLICATION FOR MEMBERSHIP

Fill out and return to

LIBRARY & INFORMATION TECHNOLOGY ASSOCIATION
AMERICAN LIBRARY ASSOCIATION
50 East Huron Street, Chicago, Illinois 60611

NAME (PLEASE PRINT)

MAILING ADDRESS

CITY STATE ZIP

INSTITUTION

POSITION

CITY STATE ZIP

☐ Yes, I want to join ALA and LITA for $55 ($30 special half price rate for new ALA members* and $25 for LITA membership).

☐ Yes, I want to add LITA to my current ALA membership for $25. My ALA membership number is _____
(check membership card or *American Libraries* label for number).

☐ Yes, renew my membership in ALA and LITA for $85 ($60 for ALA membership, $25 for LITA membership).

☐ Enclosed is my check made payable to the American Library Association.

☐ Please charge my membership dues to
☐ Visa ☐ Master Charge

EXPIRATION DATE CARD NUMBER

SIGNATURE

*ALA Membership benefits include *American Libraries*, substantially reduced rates at ALA conferences, discounts on ALA monographs, eligibility for ALA offices and committee appointments, an excellent group insurance plan, and the backing of the strongest unified voice for libraries and librarians in the world.

We've taken book ordering out of the Dark Ages.

BaTaSYSTEMS™ is Baker & Taylor's newest generation of electronic book ordering services. It's especially designed to work with existing computer hardware, with built in flexibility that allows you to match the level of service to your library's unique needs.

Whichever service level you choose, you'll save time, reduce paperwork and speed book acquisitions—all at a lower cost. For example:

ORDER allows you to order books through your personal computer, using a modem and regular telephone lines. Just enter the ISBNs and the following day you'll receive electronic confirmation from which you can print order slips. All calls are toll free. You also save the cost and delay of postal delivery.

Or you can choose SEARCH AND ORDER. In addition to electronic ordering, this service gives you quick access to Baker & Taylor's diverse and comprehensive database of over 800,000 title records. It's your single source for virtually all the titles published or distributed in the United States. And you eliminate manual searching and purchase order typing.

Finally, BaTaSYSTEMS ACQUISITIONS offers on-line access to our database and electronic ordering plus a complete software package with fund accounting and full reporting functions.

These advanced service technologies are typical of how Baker & Taylor stays in step with the times, building on our experience to bring you the latest in library services.

BaTaSYSTEMS. It's nothing less than a renaissance in book acquisitions. Write or phone today for more information.

EXPERIENCE YOU CAN DEPEND ON
BAKER & TAYLOR
a GRACE company

Eastern Division, 50 Kirby Avenue, Somerville, NJ 08876 (201) 722-8000
Southern Division, Mt. Olive Road, Commerce, GA 30599 (404) 335-5000
Midwestern Division, 501 S. Gladiolus Street, Momence, IL 60954 (815) 472-2444
Western Division, 380 Edison Way, Reno, NV 89564 (702) 786-6700

Reports and Working Papers

SISAC: Standardized Formats for Serials

Patricia E. Sabosik

This paper provides the library community with a status report on the voluntary standardized formats for electronic transmission of serial information being developed by a U.S.-based industry group, SISAC. The focus of the paper will be the workings of two SISAC committees, the Serials Orders and Claims Format and the Serial Identification Code, and will cover the following areas: background information on SISAC, standards development in the U.S., a discussion of the two proposed formats, computerized orders and claims and the identification code, and concluding statements on the benefits of standards.

BACKGROUND INFORMATION

SISAC stands for the Serials Industry Systems Advisory Committee. The committee was formed in 1982 at the invitation of the Book Industry Systems Advisory Committee (BISAC) for the purpose of developing standardized computer-to-computer formats for the serials industry. BISAC at this time had already spent eight years developing a standardized computer-to-computer format for book ordering and a modification of this format for serials forms the basis of SISAC's proposed standard for ordering and claiming serials. Early participants in the Serials Committee were Richard Rowe, president of Faxon and nominated to chair SISAC; Wendy Reidel, head, national serials data program at the Library of Congress and chair of subcommittee no. 1, the identification code; Asha Capoor, director of technical services at Baker & Taylor and presently chair of subcommittee no. 2, order and claim format; Patricia Sabosik, then director of marketing for H. W. Wilson and chair of subcommittee no. 3, membership and publicity; and Sandra K. Paul of SKP Associates and managing agent of the Book Industry Study Group, Inc.

BISAC and SISAC are committees of the Book Industry Study Group (BISG). BISG is a voluntary membership organization formed to assist the U.S. publishing industry in the exchange of ideas between members and to promote and support publishing research. Publishers, retailers, wholesalers, and manufacturers form the majority of membership in BISG. There are also members categories for nonprofit institutions, libraries, university presses, and individuals. The membership body today represents all areas of the U.S. book industry and is beginning to branch out into the serials industry.

The membership of SISAC comprises librarians, publishers, database producers, retailers, subscription agents, and information specialists, each with specific concerns relating to the efficient processing and transmission of serial information. The serials industry is SISAC's common link, and the committee is an industry forum for addressing serial industry concerns.

SISAC's planning committee identified three areas in which SISAC can make a contribution to the efficiency and effectiveness of serial publishers, database producers, libraries, librarians, subscription agents, and other information users and providers:

1. Standardization of formats to allow for computer-to-computer ordering and claiming of serials. (More than seventy publishers can accept the BISAC tape-to-

Patricia E. Sabosik is vice chair of the Serials Industry Systems Advisory Committee (SISAC) and editor and publisher of *Choice*, the review journal of the Association of College and Research Libraries.

tape order format, which is reported to reduce title research and key strokes, thereby saving time in processing such orders as well as reducing errors in shipments.)

2. Automation of library check-in, circulation, and control through standardized machine-readable coding to identify specific issues of serials.

3. Standardization of machine-readable coding of articles to facilitate (a) the payment of royalties; and (b) the identification and retrieval of articles in full-text and bibliographic databases.

In the three years since SISAC has been working in these areas, two proposed standardized formats have been developed. Addressing the first concern is the Computerized Serials Orders, Claims, Cancellations, and Acknowledgments draft standard; the second and third concerns are addressed in the Serial Issue/Article Identifier Code. The two SISAC committees responsible for developing these formats also have committee status in the National Information Standards Organization, NISO (Z39). SISAC's role as an industry group in the standards arena is to develop voluntary standardized formats for serials to present to NISO for consideration as a national U.S. standard.

NISO (Z39) is the committee of the American National Standards Institute that coordinates standards activities pertaining to libraries, information science, and related publishing practices on a national level. NISO participates in the International Standards Organization (ISO) in the preparation of standards on the international level.

STANDARDS DEVELOPMENT IN THE US

Standards development in the United States is a voluntary activity. As such, it is difficult to determine the actual cost of bringing a standardized format to publication and adoption as a national standard. Current thinking is that it is more expensive not to have a standard than it is to develop one, and, since standards are a voluntary activity, there are no direct recoverable research and development (R & D) costs a firm has to deal with prior to implementation. R & D costs are borne directly by the volunteers of the standards committee giving their free and professional time to work on standards development and by businesses, libraries, and other agencies and groups that provide release time and expense compensation for staff involved in the standards process.

The NISO (Z39) standards for publishing practices and the ANSI X3 standards for communication as well as other ANSI formats are all standards voluntary for adoption. There is no legal force requiring a firm to integrate a standard into its operation once it is developed. Pressures to implement standardized formats come from the user communities, and this pressure on publishers and information providers will increase as more standards are developed that deal with electronic access to, and distribution of, information. Publishers may also find their information excluded from certain distribution systems if they do not keep up with standardized formats of communication and information exchange.

Standards cost money to develop and to implement. How much money is a function of the length of time it takes to develop the standard (over ten years for the BISAC variable-length order format); the amount and complexity of the software required to support the standard; and the number of groups initially affected by the standard. In addition to the initial cost to develop a standard there are ongoing costs of operation and possibly software enhancement as the user community molds the use of the standard to actual needs through implementation and attempts to keep current with information technology. American standards go through a revision process every five years. Though standards development is a relatively small activity of the information industry, the effects on the manipulation and distribution of data can be far-reaching.

COMPUTERIZED SERIALS ORDERS AND CLAIMS

In December 1985 NISO released for public review a draft version of American National Standard Z39.55-198X Computerized Serials Orders, Claims, Cancellations, and Acknowledgments. The draft standard was originally developed by a

committee of SISAC and is now being proposed as a United States standard.

The proposed standard covers the functions of computerized ordering and claiming of serials and the cancellation and acknowledgment of serials orders. The standard takes a triangle approach to transmitting serial information between publishers, vendors, and libraries with all transmissions in the same format (see figure 1). The electronic transmission of this serial information is a function of library, vendor, and publisher automation and the types of computer facilities each one uses.

The SISAC order and claim format, as previously mentioned, is based on the BISAC format for computer-to-computer book ordering specified in the American National Standard ANSI Z39.2-1979, Bibliographic Information Interchange on Magnetic Tape. The BISAC format is based on the Library of Congress's Machine-Readable Cataloging Record, the MARC format. Both record structures are similar to the MARC record consisting of a fixed-field header record and a directory of variable-length data fields drawing on the flexibility and success of the MARC record to handle large and/or complex record types in automated systems. In library automation systems, both the BISAC format for books and the SISAC format for serials require a conversion program to bridge the ordering format to specific fields in the MARC record.

It is a goal of the BISAC and SISAC formats to be able to cover all material that is published. Between the two formats there will be a few gaps and some overlap; 100 percent coverage was not an objective during development, but rather the committees sought to develop computer formats that are practical and operational. The scope of the material covered by the SISAC format is primarily any item that has an ISSN.

The SISAC order and claim record is composed of four areas (see figure 2):

1. A *leader* consisting of twenty-four characters.
2. A *directory* consisting of a three-digit tag for each data field and its starting character position relative to the first data field.
3. *Data fields* of varying length, each separated by a field terminator. (The data in these fields may be divided into subfields.)
4. A *record terminator* (*R/T*) at the end of the last data field.

This format is intended to be a carrier format for exchange purposes. It does not stipulate the form, content, or record struc-

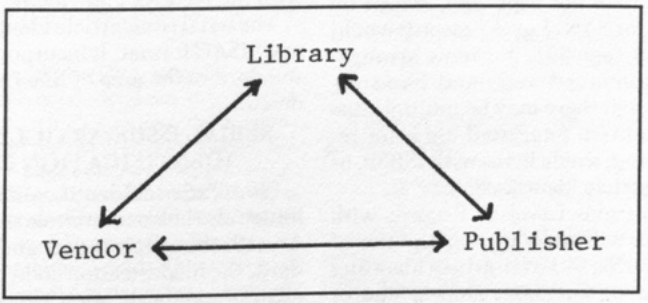


Fig. 1.

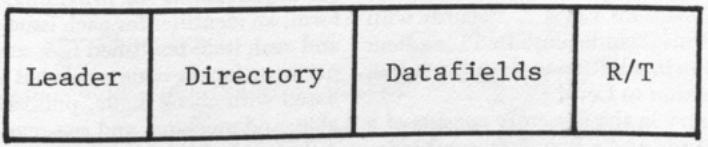


Fig. 2

ture of the data *within* individual systems. This is where bridge programs come into play.

The *leader* contains data related to the structure of the record and is a required part of the SISAC format. The *leader* data also relate to the type of record, record status, and the degree of completeness of the record. The status code indicates the relationship of the serial record to the vendor's file, the type of transaction being electronically transmitted, and can have one of the following status codes:

A = Acknowledgment
C = Claim
D = Cancel
M = Modify
O = One Time Order
R = Renewal
S = Ongoing Order
T = Transfer
V = Varies

The record-level code, also in the leader, indicates the degree of complexity of the record. Three values (A, B, or C) are covered in the standard and a description of each follows:

A = Acquisitions Level 1. This is the minimum information required to transmit a record, e.g., only the SANs and ISSNs, ISBNs, ISSNs, or serial issue/article identifier are transmitted. Authors, titles, addresses, etc., are not expressed, except by ISBN, ISSN, or SAN. Level 1 records would include only tags 001, business arrangements, customer address, and items ordered, although there may be multiple line items transmitted. Suggested use is for renewal ordering, single items with ISBNs, or serial issue/article identifiers.

B = Acquisitions Level 2. Records with Level 2 status will include a combination of SANs and ISBNs, or serial/article identifier with vendor or customer ship-to and/or bill-to amplification, e.g., business arrangements, vendor and customer addresses, and line items ordered.

C = Acquisitions Level 3. Records with Level 3 status include amplified Line Item Information in addition to the information included either in Level 1 or 2.

Each entry in the directory consists of a three-digit tag and a five-digit number indicating the starting character position of the field referenced by the tag. The directory is a required part of the format and is followed by the data fields. The data fields are divided into subfields, and much of the data identified in this area of the proposed format is optional. Data fields cover the following information and the number of subfields identified in each area are displayed after the title of the field.

Tag Data Fields

010 Business Arrangements - 21
020 Vendor SAN Details - 11
030 Customer SAN Details - 27
040 Line Items: ISSN, ISBN, Quantity, etc. - 34
050 Additional Line Item Information - 31
060 Order Acknowledgments/Order Cancellation Acknowledgments - 9
070 Claim Acknowledgments - 4

Electronic addresses for vendors and customers are included as subfields in the vendor and customer address areas. As the computer-to-computer transmissions environment expands, a need will arise for designation of an electronic address to which information may be sent. The form of the address is not known at this stage of development, but it may contain information about which public packet-switching network the vendor or customer subscribes to or some other way to describe the method by which the sender will communicate with the receiver and vice versa.

The serial issue/article identifier, the second SISAC format, is incorporated into this standard in the area of line items being ordered.

SERIAL ISSUE/ARTICLE LEVEL IDENTIFICATION CODE

SISAC's Serial Identification Code Committee also holds committee status in NISO. As with the serials orders and claims standard, the identification code builds on preexisting standards, the ISSN and the numbering schemes used for citing serial chronology and enumeration, and defines the requirements for providing, in coded form, an identifier for each issue of a serial and each item published in a serial. Early priorities for the code were that it be ISSN based with check digits, publisher assignable, and machine- and eye-readable. The code uses and extends the ISSN for serial title control.

The NISO and SISAC joint committee developed the numbering scheme for the identification code. SISAC then took the scheme one step further and developed a machine-readable format. This format uses bar code symbology for data transmission, specifically Code 128.

The serial issue identifier is made up of elements that relate to various serial components. The two main areas covered in the issue identifier are enumeration and chronology. Enumeration identifies one specific issue of a serial title, and, to construct the enumeration portion of the code, the volume and issue information that appear on the serial is used. Up to four levels of enumeration can be used in the issue identification process plus a level for identifying supplements and indexes. The chronology identifies a specific date for one specific issue of a serial title. All dates are recorded numerically in the format YYYYMMDD.

Figure 3 shows the maximum length of the code including the article identification information. It is anticipated that the first use of this code will be for issue identification. In this case only the ISSN, chronology, enumeration, and check digit will be used.

Check digits are built into the serial identification coding scheme at different levels of the code. The following explanation is taken from draft version four of the proposed standard:

> The check digit for the Serial-Article Identifier is calculated by applying the modulus 37 technique to the characters which compose the human-readable format of the ISSN, the chronology, the enumeration, the page number, and, if present, the title code. When the Serial Article Identifier Check Digit is calculated, the Serial Issue Identifier Check Digit and its preceding hyphen are dropped from the human-readable format and not used to calculate the Serial Article Identifier Check Digit. Every human-readable format will have one, and only one, check digit. This check digit will always be the last character, and it will always be preceded by a hyphen.

The serial article identifier is an extension of the issue-level identifier and consists of the previously defined data plus two additional data elements, a four-character (or less) alphabet code taken from the article title and the initial page number within the issue on which the article begins. The title code can be constructed by manual or computer-based means following algorithms set down in the draft.

The following examples help to illustrate the length and format options of the Serial Issue and Article Identifier.

1. Fixed Length Human-readable Format.
   See figure 3.
2. Variable Length Human-readable Format.
   1234-5678(1985)-7 = 1985 Annual
   1234-5678/44:A:1-9 = v.44,pt.A,no.1 (No Chronology)
   5678-1234(198509*)-3 = 1985 April Index
3. Fixed Length Machine-Scannable Format
   12345678198509050044 00A0100090144 ATCIT
   This format requires no punctuation because each data element falls into a fixed location.
4. Variable Length Machine-Scannable Format.
   12345678198509 = 1985 September
   The punctuation used in this format will be determined by the method of electronic data transmission. After transmission, the data will be transcribed back into the human-readable format.

These examples are also taken from draft version four.

BENEFITS OF STANDARDS

In a service economy and an information society, standards ease access to information; they break down barriers. From a *processing* perspective users can handle

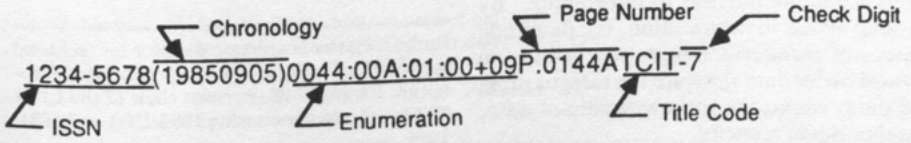


Fig. 3. Serials Issue/Article Level Identification Code.

data more effectively. For example, the work that SISAC and NISO have done in joint committee to develop a standardized order and claim format and a numeric code for identifying serials at the issue and article level. The proposed numbering scheme for the ID code is based on the ISSN, the order and claim format is based on the BISAC format for electronic ordering. The implication for libraries will be easier handling of serials, particularly for ordering and at check-in.

From a *marketing* perspective, two major benefits of standards, particularly those related to electronic publishing, are (1) shorter distribution channels for information producers and consumers; and (2) access to world markets. Information as a commodity spurred the development of different channels of distribution and new types of channel participants, e.g., online vendors, who move the information commodity from producer to consumer. Standardized formats facilitate this information exchange. Frequently there are value-added services attached to the information provided by the distributor.

Some distribution considerations in the transmission of information are (1) information is an intangible; (2) information is characterized by a lack of inventory; (3) production and consumption occur almost simultaneously; therefore, very distinct channels of distribution exist; and (4) information is perishable due to the nonstorage aspect. Without a physical flow or the need for inventory or storage, multiple resellers do not play a major role in the distribution of information as they do in the distribution of a physical product. Standards play a role in eliminating resellers or intermediaries in a distribution channel.

The second marketing benefit, access to world markets, deals with the issue of transborder data flow. Implementation of international standards eliminates or eases cultural, language, physical, geographic, and economic barriers to information. By easing access to information, the downside issues of standards integration leading to transborder data flow are the safeguarding of data, copyright and ownership of data, and national security.

SISAC BETA TEST

In spring 1986 SISAC, with the cooperation of a selected group of journal publishers, libraries, and system vendors will test the operation of the identification code for serials at the issue level in the bar code format. The issue-level format will be tested in serial check-in and control systems, and the participants in the test will follow the triangular approach to data transmission previously mentioned. Mary Ellen Clapper of the Faxon Company is the SISAC beta test coordinator, and questions concerning the test should be directed to her at Faxon. The serials order and claim format, ANS Z39.55-198X, will also be tested; a test schedule has not been established at the time of writing.

Copies of both SISAC proposed formats are available, on request, from NISO. Contact Patricia Harris, Executive Director, National Information Standards Organization (Z39), National Bureau of Standards, Administration 101, Library E-106, Gaithersburg, MD 20899; (301) 921-3241.

Inquiries for the Serials Industry Systems Advisory Committee should be addressed to SISAC, 160 Fifth Ave., New York, NY 10010; (212) 929-1393. ■■

Needed Technical Standards: Current Status

Ruth C. Carter

During his year as president of LITA, Ken Dowlin's theme was technical standards. With that as a priority, Dowlin, in June 1983, charged the LITA Technical Standards for Library Automation Committee (TESLA) with identifying those technical standards that did not exist but were needed. During the next two years,

Ruth C. Carter is assistant director for technical services, University of Pittsburgh Libraries, Pittsburgh, PA 15260. Carter was chair of the LITA TESLA Committee during 1983/1984 and 1984/1985.

one of the major activities of TESLA was the development and refinement of a list of needed technical standards relevant to library automation.

This communication reviews the work on identification of the needed standards, describes recent action initiated in several instances, and comments on future possibilities for this activity.

The first step taken to fulfill the task of identifying needed technical standards was an informal survey conducted by TESLA. Members of TESLA and other interested parties were contacted and asked to list any technical standards that were needed. Returns from the survey and further discussions at the TESLA meetings at the ALA Midwinter Meeting in January 1984 led to the development of an initial working list. This list and a more extensive description of activity leading to the list were described in an article by Carter in early 1984.[1] Each suggested technical standard was listed along with a recommended action and/or comment.

At each of the three succeeding ALA meetings (June 1984, January 1985, and July 1985) a considerable segment of the TESLA meetings was devoted to a refinement of the list of suggested standards.

As of January 1, 1985, the list of suggested needed technical standards is shown in table 1.

The list was reviewed at the TESLA meetings at the July 1985 Annual Conference. The intent of that series of meetings

Table 1. Suggested Needed Technical Standards as of January 8, 1985

	Suggested Technical Standard	Status/Action to Take
Display/ Retrieval	Basic display for online catalogs	Monitor
	Online catalog searching standards	In progress—Z39 G
Data Elements/ Formats	Common bibliographic database	In progress
	Collection locations	Partially in progress—with Z39 V
	Patron record standard data elements	Monitor/may not be possible
	Local news index records	Continue to investigate
	Format for machine-readable classification schedules	Monitor
	Minimal-level catalog record data elements	Monitor
	Circulation record form and content	Monitor—vendors want
	Index construction algorithms (normalization of bibliographic data)	Needs active investigation and introductory article
	Machine-readable item identifier codes (OCR & bar codes)	Monitor
	Common character code for East Asian machine readable vernacular characters	Monitor
	Data elements for information referral	Monitor
	Data elements/format for ready reference files	Monitor
Linking/ Communications	Common communications interface to link turnkey systems—circulation, acquisitions, etc.	Discussions underway
	Circulation system record—link to bibliographic record	Monitor
	Authority record to bibliographic record link	Monitor
Miscellaneous	Security procedures to protect online data	Not for a standard
	Promote videotext screen standards for home access to public catalogs	Matrix
	Standard for graphic storage & display	Matrix
	Performance data for systems	Initiate action

was not to perpetuate the list as an entity, but rather to identify those items on the list that could be moved to an action status. Consequently, it was agreed that outlines of the statement of issues for each of the following four topics would be developed. Those topics are

1. performance data for systems;
2. circulation record form and content;
3. common character code for East Asian machine readable vernacular characters; and
4. index construction algorithms.

Progress should be made, then, in the near future on defining the issues and determining appropriate steps to be taken to initiate work on developing standards appropriate to the four items listed above. At the same time, TESLA expects to continue in its role of assisting in the development and promotion of technical standards for library automation in the broad context of library automation.

TESLA can be expected to fulfill its mission in partnership with the various standards developing agencies. These include the American National Standards Institute X3-Information Processing Systems Committee (X3) and the National Information Standards Organization (Z39) (NISO [Z39]). It is worth noting that in April 1985 NISO (Z39) sponsored a one-day futures conference. Using the nominal group technique where all participants suggested areas for future activities for NISO and then developed priorities through a series of group and individual rankings, the brain storming by attendees resulted in a list of more than sixty-five topics for future action.

On close examination a few major topics emerged, several of which were in the area of needed technical standards. In fact, the number one item was linking systems—at all levels and of all types. Because this topic has moved to an action status through the efforts of NISO and others, including vendor groups, linking systems is not included in the most recent list of areas targeted for background work by TESLA members. The NISO conference did, however, reinforce the needs for TESLA activities both as an identifier of standards that are needed and as a vehicle for providing standards publicity and heightening general awareness of existing standards.

REFERENCE

1. Ruth C. Carter, "Identifying Needed Technical Standards: The LITA TESLA Committee at Work," *Library Hi-Tech* 5:37–40, 1984. ■■

News and Announcements

AMIGOS, UTLAS to Offer Reciprocal Services

AMIGOS Bibliographic Council and UTLAS have announced an agreement by which AMIGOS will offer UTLAS' authority control services and UTLAS will provide its Canadian customers with access to AMIGOS' Collection Analysis Service.

Effective immediately, UTLAS authority control will be available as part of AMIGOS' retrospective conversion and archival tape processing services. UTLAS' services provide up-to-date authority control for name, title, series, and subject headings.

AMIGOS has offered authority control for name headings based on OCLC's AACR2 headings "flip" since 1982. "We began to need a more comprehensive authority control capability," said Louella V. Wetherbee, executive director of AMIGOS. "The UTLAS agreement allows us to offer authority control right away, without additional programming effort."

UTLAS now will offer AMIGOS' Collection Analysis Service to its Canadian customer base. The AMIGOS service uses library tape files to provide a comprehensive analysis of holdings, including identification of collection strengths and weaknesses, unique titles, duplicate holdings, and collection growth trends. ■■

EBSCO and Sydney Announce Interface

An interface that will substantially enhance serials management and order processing for libraries has been announced by EBSCO Subscription Services, a worldwide agency headquartered in Birmingham, Alabama, and Sydney Development Corporation, a developer of automated library systems headquartered in Vancouver, British Columbia, Canada. The new interface, between EBSCONET and the Sydney Micro Library System, provides online transmission of claims and orders for EBSCO subscribers and Sydney Micro Library System users.

The Sydney Micro Library System offers an integrated library automation package where bibliographic and status information are shared by all modules. This system automates the inquiry, cataloging, circulation, acquisitions, and serials-control functions for libraries. The system features sophisticated inquiry and can accommodate MARC-formatted data elements. The system is modular and available in single user or multiuser versions for microcomputer-based local area networks.

EBSCO Online Subscription Service offers libraries serials management through online access to more than 165,000 serial titles. As an extension of its service, EBSCO has as a major objective the provision of machine-readable subscription information to its users.

The new interface makes use of Sydney's Serials Control Module to transmit claims and orders online to EBSCO. Sydney's Serials Control Module generates various management reports, processes invoices and renewal orders, updates subscription and invoice files, and provides check-in, routing, and claims and order processing.

The major component of the interface is the online transmission of claims and orders from a Sydney user's IBM PC or compatible microcomputer to EBSCO's IBM mainframes, using standard communications protocols on the UNINET X.25 packet switching communications network. Data editing for errors is performed at the time of transmission. This online procedure results in reduced errors, faster transmission, reduction in communication costs, and faster satisfaction of claims.

For more information, contact Mary Beth Vanderpoorten at EBSCO at (205) 991-1368, or David Devine at Sydney Dataproducts at (213) 479-4621. ■■

RLG Inaugurates Transatlantic Link with British Library

On October 16, 1985, the transatlantic telecommunications link between RLG and the British Library in London became operational. The link, installed last June, is now being used in an international project coordinated by the British Library to create a computer database that contains records of books, pamphlets, and ephemera printed in the English language during the eighteenth century, anywhere in the world.

The two editorial teams working on the project—the Eighteenth-Century Short Title Catalogue (ESTC)—in London and at ESTC/North America, in Baton Rouge, Louisiana, are now both entering records online into RLG's Research Libraries Information Network (RLIN). Prior to the link, the London team's records were tape loaded into the RLIN database once a month, and project editors worked on two separate files that were never quite in sync. ■■

Cornell Selects BLIS for Public Access Catalog

Cornell has selected Biblio-Techniques Library and Information System (BLIS) to serve as an online public-access catalog and integrated circulation control system for all of the university libraries. Cornell and Biblio-Techniques will undertake a six-month demonstration, planning, and training project as the first phase of a full implementation of BLIS.

The BLIS software will run on an IBM mainframe computer operated by the university. A joint development project between Cornell Computer Services and Biblio-Techniques is being undertaken to produce an interface for BLIS software to use IBM's VM/CMS operating system software.

The BLIS software was installed at Cornell in early November. The training of Cornell's Implementation Group and preparation of data-conversion specifications will be complete in the first quarter of 1986. Biblio-Techniques will provide custom data conversion services to load Cornell's RLIN database of more than one million records. These records represent approximately one-half of the titles in Cornell's collection of five million books and three and one-half million microforms.

Cornell expects to continue using the RLIN system for its cataloging and acquisitions operations for at least the next two years, before moving those activities to its local system.

The Joseph N. Pew Charitable Trust provided the funding for Cornell to undertake this library automation effort.

Cornell University is the seventh site to install BLIS software. Other BLIS customers include Columbia University, University of California at San Diego, University of Cincinnati, Metro Toronto Library, Indiana University, and Brown University. The WLN software, on which BLIS is based, is also installed at the University of Illinois, University of Missouri, SOLINET, the National Libraries of Australia, New Zealand and Singapore, and the British Library. ■■

BCI Licenses iNet 2000 Software to OCLC

OCLC has acquired a license for the software for Telecom Canada's iNet 2000 intelligent data network.

OCLC will use the iNet 2000 software in its private telecommunications network to provide intelligent gateway features to its member libraries. A date for the implementation of those features has not yet been announced.

Telecom's iNet 2000 has been available since 1982 and was implemented by Telecom Canada to provide user-friendly aid and guidance in selecting databases that address the user's need from among hundreds of databases available. The service was in trial stages from 1982 to 1985 and became a full service offering in November 1985.

Additionally, iNet 2000 helps the user in accessing online databases and in communicating that information through electronic messaging and data conferencing.

More than sixty-three hundred users are served by forty-six information service providers on the Canadian network.

OCLC will incorporate iNet 2000 features such as

- a single point of access to diverse sources of electronic-based information, including those not traditionally available through the large, centralized distributors;
- assistance in identifying the most appropriate source of electronic-based information for a particular need;
- the ability to manipulate and communicate acquired information for further dissemination;
- ease of implementation of such services as electronic mail, linkages to private networks, private networks configured using the OCLC network, and a single bill for all use of information sources on the network.

The product is marketed in the United States by BCI of McLean, Virginia, a subsidiary of Bell Canada International, the international telecommunications consulting and operations subsidiary of Bell Canada Enterprises (BCE). The software was developed by Bell Canada, BCE's telecommunications operating subsidiary and the largest member of Telecom Canada, an association of ten major telecommunications companies across Canada. ■■

NICEM Database to be Offered on CD/ROM

Access Innovations of Alburquerque, New Mexico, and SilverPlatter Information, from Wellesley, Massachusetts, have signed an agreement to develop a CD/ROM product for the NICEM database. The National Information Center for Educational Media (NICEM) database is currently available in printed publications or through online services.

By offering the NICEM database on CD/ROM, Access Innovations will offer NICEM users the ability to continually search and retrieve information from the complete database at an independent work station in their own library. The system will operate on an IBM work station with 256kb memory, keyboard, display moni-

tor, and a CD/ROM player.

SilverPlatter will supply all the technical services required to prepare the CD/ROM product. These include all software for disk preparation and search/retrieval; conversion and inversion of data for index creation; premastering, mastering, and copying of the disks. SilverPlatter is also providing comprehensive user support service and other technical assistance required by the information provider, Access Innovations.

The CD/ROM product will incorporate sophisticated search and retrieval software developed by SilverPlatter, which will allow NICEM users to search the database through key words, descriptors, or by words in any field of data. This CD/ROM product will allow simplified user research into the availability of audiovisual materials that include videotapes, films, filmstrips, and audiocassettes.

The NICEM database on CD/ROM will be available for purchase by users of the database in April 1986. It will cost $800 annually for a subscription. Subscribers will be educational institutions, school districts, and libraries. ■■

Personal Bibliographic Software Announces Pro-Cite

Personal Bibliographic Software has announced the new, enhanced version of their professional bibliographic system—Pro-Cite. The system was designed for the storage and retrieval of bibliographic records on the microcomputer, based on the specific needs of scholars, librarians, and researchers.

Pro-Cite provides flexible data entry work forms that allow users to create and maintain a database of records that may be printed according to the style sheet of any journal or association. Full-text searching and sorting on all fields is provided.

New features include

- all variable length fields, including variable length key word field;
- twenty-two document types for data entry, including two user-defined work forms;
- full-text searching—unlimited key word searching;
- full Boolean logic—AND, OR, NOT, parentheses, truncation, <, and >;
- sorting on all fields;
- customized bibliographic formats through a menu-driven user interface;
- deletion of duplicate records automatically;
- the ability to scan a document for in-text references and automatically produce a bibliography from all collected references.

With the Biblio-Link companion programs, users can transfer data into Pro-Cite from Dialog, BRS, OCLC, or RLIN. Biblio-Link reformats the downloaded bibliographic data from the online system and places it into the specified Pro-Cite database.

Pro-Cite and Biblio-Link were developed by Personal Bibliographic Software of Ann Arbor. The company, incorporated in 1982, markets microcomputer software for bibliographic applications to libraries, educational institutions, research institutes, and government agencies. Pro-Cite is the new, enhanced version of the firm's professional bibliographic system. Pro-Cite is available for the IBM PC, IBM PC XT, IBM PC AT, OCLC M300, and IBM compatibles. List price for Pro-Cite, $395, list price for Biblio-Link, $195 for each host system.

Personal Bibliographic Software also publishes bibliographic software for the Macintosh, Apple II, IIe, IIc, the Texas Instruments Professional Computer, and the DEC Rainbow. ■■

Information Access Company Introduces Government Publications Index on Video Disk

Information Access Company (IAC) has introduced a patron-oriented, current index on videodisc covering documents published by the U.S. Government Printing Office (GPO). Called Government Publications Index, the reference system enables a user to find citations quickly to publications issued by the GPO and is designed for use with the IAC InfoTrac system.

Government Publications Index offers a fully cumulated and integrated index of the eight most recent years of contents of the monthly catalog of the GPO. The reference

system is updated monthly with the delivery of a new videodisc containing the cumulated database together with the most recent entries. Government Publications Index is designed to reflect the researcher's needs with four alphabetized access points for each document. Users may search Government Publications Index by subject, title, author, or issuing agency. Library of Congress subject headings and subheadings are used as well as a complete system of cross-references.

In fall 1985, IAC introduced Government Publications Index, a microfilm reader system that also covers eight years of the monthly catalog of the U.S. Government Printing Office. This microfilm reference system is also updated monthly with a new, fully cumulated reel of microfilm containing the new listings.

A subscription to Government Publications Index can include up to four microcomputer access stations and printers in addition to the videodisc player, software, controller, and twelve monthly cumulated updates on videodisc. Government Publications Index is the fourth videodisc database to be available for use on the InfoTrac system. The InfoTrac database, which indexes approximately 900 general interest and business magazines and newspapers, was introduced in January 1985; LegalTrac, which indexes more than 730 law journals and newspapers, was introduced in April 1985; and a videodisc containing the full-text coverage of the *Wall Street Journal* was first demonstrated in November 1985. ■■

OCLC Opens Gateway to BRS

OCLC has announced the availability of the OCLC Gateway Service to BRS, which enables OCLC members to have access to BRS databases via dedicated OCLC terminals. The Gateway Service provides those OCLC members who are also BRS subscribers with a single point of access from which library staff can effectively use both BRS databases and the OCLC Online System. From a single OCLC dedicated-line terminal (Model 100, 105, 110, or M300 Workstation), users can access the OCLC Online Union Catalog with more than 13 million bibliographic records and 220 million holdings symbols, as well as more than 100 BRS databases containing information on a wide variety of subjects.

Once logged on to BRS, users can use BRS system commands to search the BRS databases and select paragraphs that correspond to patron needs. A BRS COST command enables users to estimate the cost of each search, thus controlling online charges.

The power of BRS' free text and keyword search system is available to users (except for terminal configuration commands) to search and retrieve data for library patrons. OCLC members who are currently BRS subscribers will not require additional training to use the Gateway Service. Billing for BRS services used over the OCLC Gateway Service is done by BRS, or through participating Network Offices, if appropriate. ■■

Bowker CD/ROM

The R. R. Bowker Company has announced it is entering into a joint venture with International Computaprint Corporation (ICC) of Fort Washington, Pennsylvania, and Online Computer Systems (Online) of Germantown, Maryland, to provide access to the Bowker databases in CD/ROM (compact disk–read only memory) format by June 1986. Among the products formatted in CD/ROM and scheduled for introduction in June 1986 are Bowker's *Books in Print* and *Ulrich's International Periodicals Directory*, both information reference standards in the library world. In addition, Bowker will unveil a number of other bibliographic CD/ROM products.

Bowker, ICC, and Online will provide libraries with complete CD/ROM services, library databases, CD/ROM and PC hardware, retrieval software, and national field support. Both ICC and online will provide preprocessing, premastering, and retrieval software to information providers.

ICC is a prepress database service company for the publishing industry. ICC pioneered the computerized typesetting industry twenty-three years ago and currently typesets in excess of three million pages per year for more than two hundred government and commercial publishing clients. ICC provides information processing services for all media, including print, electronic, microfiche, and optical disk.

Online Computer Systems will also provide turnkey hardware, installation, and maintenance to libraries nationally. Online specializes in information delivery systems in a variety of forms, with expertise in electronic publishing and library systems. Online offers a range of services (including technical support) to allow the publisher full control of system design and operation. Earlier this year the Library of Congress awarded Online a contract for the design phase of its pilot project to distribute cataloging data on CD/ROM. ■■

Recent Publications

Book Reviews

Binder, Michael B. *Videotex and Teletext: New Online Resources for Libraries.* Greenwich, Conn., and London: JAI, 1985. 160p. Foundations in Library and Information Science, V.21. ISBN: 0-89232-612-3, hardcover, $47.50.

Videotex, without the second *t*, is "an interactive electronic system in which data is transmitted from a computer network over telephone or cable television lines and is displayed on a subscriber's television or computer terminal screen." *Teletext*, with the second *t*, is "an electronic system"—not interactive—"in which printed matter is broadcast by a television station and displayed on a subscriber's television set having a decoder." Both definitions come from *Webster's Ninth New Collegiate Dictionary;* both systems come from the late 1970s and from England and Europe. Almost since their inception, teletext and videotex have been viewed as key elements in the "paperless society" and as sources of potentially enormous profits. Governments and companies have poured years of efforts and tens of millions of dollars into variations on the two systems.

Michael B. Binder, director of the Rutherford Campus Library of Fairleigh Dickinson University, was sufficiently interested in videotex and teletext to take six months leave to carry out bibliographic research into the systems. The result is a compact, readable book that succeeds in Binder's expressed attempt: "presenting a view that is grounded in fact, not predictions or marketing hype." Binder believes in the potential of one-way teletext and two-way videotex, but only occasionally lets that belief color his presentation of the troubled history of both systems. The resulting book is neither the most comprehensive nor least expensive introduction to videotex/teletext, but does appear to be the first American monograph specifically incorporating the history of library involvement in videotex/teletext.

The first chapter defines videotex and teletext and relates them to other online information systems. The second and third chapters discuss the fundamentals of videotex and teletext respectively. Four following chapters review developments in Britain, France, Canada, and the United States. The final chapter concerns videotex and teletext use by libraries. An extensive bibliography, author index, and subject index complete the book.

Binder writes clearly and engagingly. The book is meant to be read rather than simply referred to. He incorporates the results of extensive research without losing readers in a morass of quotations. He defines terms as they appear and keeps the discussion interesting even as it becomes (necessarily) heavily laden with technical matters and special terminology.

The book is not flawless. A reader going from beginning to end should be able to deal with the terminology, but the lack of a glossary makes it much more difficult to refer back to elements of the book. The bibliography appears to be current through 1984, but lacks an entry for NAPLPS, the North American Presentation Level Protocol Syntax (ANSI X3.110-1983), even though NAPLPS is mentioned several times in the text.

Binder hints at including CompuServe, The Source, DIALOG, and others in the realm of videotex/teletext and once compares a videotex network to the Linked Systems Project. Neither the inclusion nor the comparison is convincing, and both seem to arise from the difficulty of reading success into any videotex/teletext ventures.

The price seems a bit steep for a relatively small book, but the production does seem professional in all respects (save the missing glossary).

This is an honest book and must have been difficult for Binder to write. He wants to put a positive face on the history he so fairly reports, but the best he can do is to say that "while videotex and teletext have not been unqualified successes, neither have they been total failures." On the evidence of this book, the truth lies much closer to the latter than to the former. Home videotex/teletext have much in common with home computing: they seem to be answers in search of questions.

Binder deliberately excludes social issues from his discussion. Given the level of impact videotex/teletext appear to have, that's a reasonable exclusion. He does provide a clear, well-researched, intelligently organized picture of videotex and teletext.

I came to the book skeptical of videotex and teletext but with little real understanding of either system. After finishing it, I remain skeptical but feel much better informed on the topic. Any librarian interested in videotex and teletext should find the book informative and useful.—*Walt Crawford, The Research Libraries Group, Inc., Stanford, California.*

Dodd, Sue A., and Ann M. Sandberg-Fox. *Cataloging Microcomputer Files: A Manual of Interpretation for AACR2.* Chicago and London: American Library Assn., 1985. 272p. ISBN: 0-8389-0401-7, hardcover, $37.50.

Over the last several years, microcomputer software has begun to invade small and large libraries in waves. Libraries that were making any effort whatsoever to provide their patrons with such materials for on-site or at-home use were presented with a bewildering array of items to which they needed to provide access; even libraries trying to avoid the issue found peculiar new items tucked away in pockets in the backs of books. Catalogers attempting to provide bibliographic records for such items were confronted with the necessity of trying to follow a cataloging code that had already been outdistanced by a changing technology, since the relevant chapter 9 of AACR2 focuses chiefly on large data files designed to run on mainframe computers. At the time of the code's publication in the late seventies, the microcomputer revolution had yet to occur.

Other than some local sets of guidelines or rather broad interpretations of chapter 9 by individual catalogers, no further "official" help was to appear on the scene until ALA's publication in 1984 of the rather brief *Guidelines for Using AACR2 Chapter 9 for Cataloging Microcomputer Software* (Chicago: American Library Assn., Resources and Technical Services Division, Cataloging and Classification Section, Committee on Cataloging: Description and Access). Helpful as this work was to the desperate cataloger, however, its usefulness was limited both by its brevity and by its lack of coverage of all the rules, background information necessary in understanding the format in order to describe it, and such topics as choice of entry, subject access, and classification. Now, with the recent publication of *Cataloging Microcomputer Files*, the individual cataloger for the first time has a thorough and clear explanation of how to apply the existing code to microcomputer software. In addition—and perhaps even more important—the work includes invaluable historical, technical, and theoretical background information in very readable form and accompanied by numerous illustrations. In fact, the first four chapters of the book—one-fourth the entire text—is devoted to providing the cataloger with the appropriate context within which to begin gaining an understanding of the format. Such a context is vital for the librarian unfamiliar with the conventions of the microcomputer and its associated software yet faced with the necessity of cataloging that software.

The central portion of the work takes the reader step-by-step through the rules in the order in which they are presented in chapter 9 of AACR2. In virtually every case, rules are cited in full, followed by sections clearly labeled as "interpretation" or "application." Where appropriate, references are made to the ALA *Guidelines* as well.

The substance of the interpretations offered in the section on descriptive cataloging is excellent. In many instances, the work will provide hesitant catalogers with the reassurance and explanation necessary

to move beyond the rules yet continue to follow their spirit. For example, the treatment of edition statements is invaluable in the way it takes the cataloger from the rather sketchy rules in chapter 9 to a real understanding of precisely what might constitute an edition statement on an actual piece of microcomputer software. Cautions such as that against using edition statements found only in some accompanying documentation will prove especially helpful to the novice.

Catalogers have always clamored for actual examples of real cataloging to aid them in interpreting the code they are applying. The Dodd/Sandberg-Fox book here again proves its value in that it provides a number of step-by-step examples of the cataloging process as applied to actual published microcomputer files. Illustrations cover the sources of information, and each area of the record receives a separate and thorough explanation. The twelve examples included should be as helpful to experienced catalogers as to novices.

The work includes several other features worthy of note. A chapter on access points will be useful, if perhaps not the first section consulted by experienced nonbooks catalogers. The chapter on classification will be of interest only to those few institutions that classify such materials; it focuses primarily on the Dewey system. The brief but important chapter on subject cataloging could profitably be read by any librarian with an interest in subject access to software; although it offers few real solutions, it represents an excellent presentation of some of the problems. The work concludes with a glossary and a brief index.

If *Cataloging Microcomputer Files* has any flaw, it is only that it attempts to deal with a rapidly changing technology by describing it at a particular point in its development. The authors themselves caution the reader about this problem in their preface, and they are perfectly correct. Yet to the cataloger attempting to provide useful bibliographic access to microcomputer software, the work under review should prove indispensible. Not only is it well written, but it also reflects a thorough understanding both of cataloging principles and the nature of the format.—*David R. Thompson, Stanford University Libraries, Stanford, California.*

Humphrey, Susanne M., and John Melloni Biagio. *Databases: A Primer for Retrieving Information by Computer.* Englewood Cliffs, N.J.: Prentice-Hall, 1986. 384p. ISBN: 0-8359-1319-8, hardcover, $25.95.

In the preface of *Databases,* the authors state, "the purpose of this book is to help those individuals, whether students or professionals, scientists or corporate executives, in their role of information consumer, so that they can effectively interact with today's rapidly emerging computerized retrieval resources. . . . with a solid grasp of basic retrieval techniques, one can be assured of success in getting information at will, instantly, and with great satisfaction." Further on in the introduction it is stated, "in writing this book, we felt that using the conventions and formats of a particular retrieval system [e.g. BRS or DIALOG] would be too confining. . . . Except for Chapter 1, which uses examples of interactions with selected retrieval services to highlight the utility of computerized database retrieval, we forego the authenticity of 'real' but rather complicated computer inputs and outputs in favor of examples that are directly to the point of the topic under discussion. The examples . . . are a composite of the real features and capabilities." The authors thus essay the extraordinary task of teaching end users the principles of database searching without ever giving details of or discussing any of the extant retrieval systems. This book is an embarrassing example of librarians trying to teach what librarians believe users ought to want to know, rather than what the users themselves are interested in.

The authors begin, after the introductory chapter containing the only authentic search histories and records to be found, by describing manual files and then comparing them with machine readable files, using school records as an example, and introducing the concepts of indexes and tables. Following is a torturous chapter on logical operators, in which Venn diagrams are depicted not as the familiar intersecting cir-

cles, but as overlapping rectangles, some of which are virtually indecipherable. Users requiring so detailed a grasp of Boolean algebra (and few ever will) would be better served by a text on the subject, such as Harold N. Lee's delightfully lucid and witty *Symbolic Logic: An Introductory Textbook for Non-Mathematicians* (New York: Random House, 1961). Subsequent chapters introduce the concept of the search statement, controlled vocabulary, free text searching, displaying output, and the like. Throughout the main section of the book, all examples reflect "retrieval from a file consisting of just two records."

In the last four chapters the authors discuss the mechanics of retrieval from a remote host, the concept of user friendliness, microcomputer use in searching, and finally, selecting a retrieval service. These chapters continue the pattern of generality previously established: they do not name specific databases or producers, telecommunications networks, microcomputers, terminals, or retrieval systems. The last chapter lists thirty-eight different criteria for selecting a retrieval service, but does not mention how well the various vendors meet them. When, finally, the retrieval services are listed by name in appendix A, the information provided is too sketchy for the reader to apply any criteria beyond database subject coverage. The reader comes away without any real information about which service will meet his or her needs, and without ever having been exposed to a typical "real life" search and result. Other oddities in the appendixes are a list of membership organizations that includes such entities as ACM and NFAIS but does not mention online users' groups and a list of ALA-accredited library schools. The thirty-eight–page list of readings (appendix C) would have profited from a subject arrangement rather than an alphabetic-by-author approach.

This book is both ill conceived and misnamed. A primer, according to the *OED*, is "a small introductory book on any subject; that which serves as a first means of instruction." This book is not that; rather, it is a generalized tract that presents in an oversimplified way a theoretical, rather than a practical, approach to searching. This reviewer cannot imagine an end user who would be willing to wade through 384 pages of a text that never even illustrates a successful log-on procedure, much less a real search. In my experience, end users tend to be "hands on" people; they already are familiar with data processing or word processing concepts and are comfortable with automation. They do not want to know "something about" searching—they want to know how to search. Such potential end users would be better advised to consult a text based on a practical approach, such as Barbara Newlin's excellent *Answers Online: Your Guide to Informational Data Bases* (Berkeley, Calif.: Osborne McGraw-Hill, 1985).

This book is unlikely to find its place as a library school text, either. In addition to the problems outlined above, the book displays a certain inattention to conceptual detail. For example, reference is repeatedly made to "records stored in the computer"; when, of course, records are stored not in the CPU, but on mag tape, disk drives, or the like. Similarly, confusion between program and hardware abounds, as in "the computer may establish certain default search keys." Those who teach the use of online systems in library school will find a more appropriate textbook in *Effective Online Searching: A Basic Text*, by Christine L. Borgman, Dinen Moghdam, and Patti K. Corbett (New York: Dekker, 1984). *Databases* is not recommended for end users, library school use, nor for the library collections.—*Jill B. Fatzer, Ohio State University Libraries, Columbus.*

Li, Tze-Chung. *An Introduction to Online Searching.* Westport, Conn., and London: Greenwood, 1985. 289p. Contributions to Librarianship and Information Science, 50. ISBN: 0-313-24274-7, hardcover, $27.95.

Beginning searchers and library administrators who are writing justifications for an online searching budget line will find a great deal of useful information in Li's book. The underlying premise throughout the book is that online bibliographic searching is an integral part of reference services, not necessarily a separate function

to be treated differently from any other reference operation in the library. He begins the work with a succinct review of the history of online bibliographic databases. The remainder of the first section examines various aspects of online bibliographic searching from the viewpoints of both the searcher and the library administrator responsible for implementing and then evaluating the service. It is useful to have all of these issues summarized and in one place.

Li states that the heart of the book is its second part, which serves as a training course for searchers and compares DIALOG, BRS, and SDC. However, the first section will have more meaning for many. The beginning searcher can use it to obtain an overview not only of the search process but also of some important issues to be considered in conjunction with online search services. For example, he delves into the reasons for and correct timing of an online search in the reference interview continuum, and he cites at least two studies in which the cost of online searching was determined to be lower than manual searching.

The discussion of the items to be considered in the initial investment for online services assumes any library planning to add the service is planning to do so by purchasing a terminal dedicated to searching. Site preparation for this terminal is mentioned, but because it may prove to be the most costly part of an installation, it would have been nice to read estimates of the cost of adding adequate wiring and telephone lines. There are alternatives to purchasing a terminal exclusively for searching. Some of the start-up costs can be avoided if bibliographic searching is added to the uses of an existing terminal. For example, some smaller libraries may wish to subscribe through OCLC, and others may be using a nonbibliographic database such as WESTLAW and choose to add the service of this terminal. Technically any terminal with modem and compatible communications package may add the service to its repertoire. This practice will eliminate some start-up costs but may create operational problems for the service. Nevertheless, a small library might wish to pursue the option; in fact many use the same terminal for many operations including searching. Leasing agreements are available for all the required equipment including the phone line for those libraries who choose not to purchase the equipment. The cost of leasing is much higher than that stated in the book—there are examples of total costs as high as $350 per month. If one believes searching to be part of basic reference services then whatever steps are needed to include the service will have to be implemented, and these options can include alternatives to purchasing. One additional cost not mentioned anywhere in the book is the cost of a maintenance contract on the equipment. This cost can be high and must be investigated before purchase. In addition, the terms of the service contract should be carefully examined because a $90 "travel" charge to make a call on a "covered" piece of equipment can be devastating to a close budget.

It is also true that many libraries try to recover all or part of the costs of online searching by charging for the use of the service. The fee versus free controversy in conjunction with the service is outlined in the section on the costs of the service in the first part of the book. Arguments are summarized and specific examples of cost recovery practices cited.

The chapter on reference sources contains a good summary of the literature; everything important is included, and the entries are accurate and informative. The literature itself has grown a great deal in the past ten years, and the division in the chapter by types of work (handbooks, directories, journals, etc.) will help determine which should become office copies and which should be considered for shelves next to the terminal. The problem with literature on or about any technological aspect of library and information studies is that the situation itself changes so rapidly that works are out-of-date by the time they appear. Among those with consistent lasting value are the articles published in journals such as *Database* or *Online Review*, which discuss the fine points of search strategies in the different subject areas. Li manages to avoid most of the technological currency traps, although one wishes there were more information on nonbiblio-

graphic databases in the work because some libraries will be able to use this type of database before the standard bibliographic ones.

The rapid growth of the literature about microcomputers and the clear dominance of IBM in the microcomputer field will render some sections obsolete before long. The detail about Scripsit as a word processing program, for example, and the information about Radio Shack micros might not be meaningful to as large an audience now as it was a few years ago. Nevertheless, the excellent introductory paragraphs in these sections will be useful for all libraries, as the type of equipment required for communications and downloading and the definition of baud rates will not change for some time. The reader should be sure not to skip the sections on a specific software because they include basic information necessary to operate any microcomputer and word processing or communications package.

If the reader searches several bibliographic and nonbibliographic databases in the course of a working day, one might wish the chapter on standardization were not buried at the end of the book. One of the most confusing parts of online searching is the difference in protocols for each system. Logging on and logging off, printing online and offline, requesting a new search, etc., become troublesome when more than a couple of systems are used. If one looks at the number of systems used in an average day by an average reference librarian, the need for standardization becomes evident. For example, there may be a circulation system, an online catalog, BRS, DIALOG, direct access to another library's online system, OCLC, RLIN, a nonbibliographic database such as WESTLAW, and an electronic bulletin board. The time spent on the intellectual part of the search and directly helping the patron is reduced by the amount of time consumed to log on and off of systems. The log-off command, for example, seems to differ for every database and system in use, and it always seems to take the longest amount of time to get off one of the systems charging for connect time. Charts help but they should not be necessary. The need for standards is increasing and should be supported wherever possible.

The second part of the book, a detailed explanation and introduction to searching on DIALOG, BRS, and SDC ORBIT is useful, accurate, and should be read before one attends a vendor course in searching. Vendor courses are generally good, but the beginner tends to forget some of the simple things such as where a space is needed, where two dots are important, or when to hit the carriage return. A review of this chapter will reinforce some of these practices and enable the learner to get more out of the course. The examples are well chosen, easy to understand, and more than a duplicate of the manual for the service. Beginners should also read carefully the section in the first part of the book, which compares searching using a thesaurus to free text searching. The latter is a temptation, especially to someone who might be moving from print to online before a complete understanding of the online structure and mechanisms is in place. The examples are especially helpful when sample searches are modified and the results included. The features selected to be expanded in the text are good ones, as they are the ones that will limit a search and help reduce the costs of online time. Again, the useful examples can be duplicated in practice.

One does wonder how useful or important the detailed examination of The Source, Compuserve, and Dow Jones is to libraries. One electronic bulletin board recently featured a discussion of the first two and the consensus was that they were the plaything of yuppies, full of bugs, difficult to get on, and very expensive. At least one group of computer literate individuals are not fans of either system. One also wonders how many libraries use them and find them to be a basic part of their reference services. Li's *Introduction to Online Searching* is a good book, worth the money. Used in conjunction with the system manual and training, it is a definite contribution to the literature and to the development of online searchers. It is well written, nicely presented, with only one typo on the introductory table of abbreviations, and well in-

dexed. It is recommended for libraries introducing online searching and for those where staff new to searching are being trained regularly.—*Ruth A. Fraley, Graduate Library for Public Affairs and Policy, State University of New York at Albany.*

Lovecy, Ian. *Automating Library Procedures: A Survivor's Handbook.* London: The Library Assn., 1984. 247p. ISBN: 0-85365-516-2, hardcover, $28 from Oryx, Phoenix.

This work presents a broad overview of library automation with emphasis on the complex series of decisions library administrators must make when contemplating automation of a range of library procedures. The author characterizes himself as a survivor of the computerization process at his institution (University of Reading, England) and states in the preface his intention to help librarians "clarify the issues which automation raises, to begin to evaluate possible solutions to their problems and finally to survive themselves."

The book is a description of alternatives on topics including the decisions on whether to automate and, if so, which functions and implementation approaches to choose; the constraints and requirements these decisions produce; and, perhaps most importantly, the impact of library automation upon staffing patterns and user satisfaction with library services. Following an introduction, the chapters are organized first by function or application, then descriptions of practical issues such as system selection, integrated systems, installation and maintenance, and troubleshooting problems. Two of the more useful chapters are those entitled "Library Staff and Library Computers" and "The System and the User."

Automating Library Procedures also contains a six-page bibliography, a glossary, a detailed index, and references at the end of each chapter. The format, coupled with the geographic emphasis on British libraries, is as significant as the content in evaluating the book's usefulness to the American audience. Just as a title published by ALA could be expected to concentrate on U.S. libraries and the American scene, this title—originally published by The Library Association, London—draws its examples primarily from libraries and programs in Great Britain. The author makes liberal use of acronyms such as SWALCAP, BLAISE, EMMA, and BLCMP, which are neither defined in a separate list nor adequately explained within the text. The use of British libraries and institutions for illustrating the experiences and pitfalls linked with the automation process is expected; however, it limits the utility for non-British audiences, and the reader is frequently confused without the contextual anchor of familiar institutions and programs.

The text also has a somewhat unavoidable problem: lack of currency. Books on library automation are frequently out-of-date in terms of emerging technologies by the time of actual publication. *Automating Library Procedures* is no exception; the chapters on functional applications highlight this problem. One of the more surprising omissions is any mention of authority control in the information retrieval chapter, which highlights online catalog issues from varied perspectives including standards such as AACR2 and the MARC format, linking library catalogs and online commercial databases, indexing, and user interface.

By contrast, the book's strengths are in the broader areas administrators must carefully analyze prior to and during computerization: automation's impact upon library staff and traditional functional divisions within an organization; communication issues; relationships between system and user; and integration of library missions, goals, and objectives with system capabilities. Lovecy concludes that one danger librarians face when automating is the temptation to organize the library around the needs of the system rather than to analyze what the system can do to improve library service.

The chapter on dealing with crisis situations is well presented. The automation planning process necessarily involves a range of factors from needs assessment and system selection through physical installation and staff training. Knowledgeable li-

brary managers and systems staff are aware of potential implementation and operational problems, but it is often difficult for the uninitiated to anticipate and plan for downtime and its attendant disruption in advance of the actual experience. This chapter reviews selected crisis issues and suggests how to either avoid or best cope with them.

The chapter on staffing implications is well organized and does an excellent job of describing the human factors that must be considered when introducing computer applications to library procedures. It stresses the importance of communication and the need to involve staff at a variety of levels throughout the implementation process and reviews typical attitudes and fears in relation to job security, loss of control, machine reliability, and job quality. Many fear their jobs will become boring or repetitive, while others question their skills in an automated environment. Staff manuals are required to explain both policies and procedures related to system operation and may need to be rewritten once sufficient online experience is gained to improve work flow and service. Much of this advice will seem apparent to the experienced systems librarian or administrator, but one must keep in mind this book's intended audience.

The final chapter, entitled "The System and the User," makes several useful points. Lovecy cautions administrators against allowing automation to depersonalize the relationship between staff and patrons and suggests that libraries be realistic with their publics regarding system capabilities in order to avoid a later assumption that the system is failing or not living up to expectations. The book does an adequate job of outlining library automation issues and excels in the less technical and more philosophical concerns such as organizational change and service enhancements.

Librarians embarking on a library automation reading program will find a broad overview of the topic here and may benefit from the single-volume treatment of a wide-ranging subject. The U.S. audience has available to it a number of books and journals that regularly cover selected library automation issues in detail. Recent issues of *ITAL*, *Library Hi Tech.*, and *Library Journal* contain articles that address the same topics either from a global viewpoint or in the context of the American library experience. Lovecy's book, by contrast, has a regional perspective that limits its usefulness in the States.—*Joan Kuklinski, Minuteman Library Network, Framingham, Massachusetts.*

Spyers-Duran, Peter, and Thomas W. Mann, Jr., eds. *Financing Information Services: Problems, Changing Approaches, and New Opportunities for Academic and Research Libraries.* (New Directions in Librarianship, no. 6). Westport, Conn. and London: Greenwood, 1985. 197p. ISBN: 0-313-24644-0, hardcover, $29.95.

In 1984 the editors organized an international conference entitled "Contemporary Issues in Academic and Research Libraries," which was held in Boulder, Colorado, and involved librarians, academic administrators, computer experts, and business executives from the United States, Canada, England, and Australia. The contributed papers for the conference have been published as two individually titled volumes. Some of the papers are included here; the rest comprise the companion volume, *Issues in Academic Librarianship: Views and Case Studies for the 1980's and 1990's.*

The dozen papers included in this volume are organized into three groups. In the first section, Sigmund Ginsburg identifies some "Problems" affecting the financing of institutions and libraries and lists emergency measures to be incorporated during times of austerity in higher education. Examined from a British perspective, the erosion of funding levels for university libraries is Samuel B. Saul's topic; he describes the dilemma of having to choose between allocating limited monies to acquisition or to staff. Librarian Ed Johnson's paper on financial planning for academic libraries takes issue with some of Ginsburg's proposed drastic retrenchment measures and instead advocates the need for librarians to undertake more goal setting and strategic planning and for academic administrators to understand library activities

better and to involve librarians more in university decision making.

In the second section, papers cover "Changing Approaches." Maurice Glicksman advocates the need for increased cooperative ventures, both within academic disciplines and among libraries. Reporting on his recent research, Paul Kantor notes that library budgets tend to be based on total holdings, rather than reflecting service costs. Dan Lester observes the declining use of library funding formulas from his survey of users of the Clapp-Jordan formula over the last twenty years. Murray S. Martin examines new needs, sources, and styles in library financial planning and stresses the importance of integrating library goal setting into university planning and of showing the library's added value to the institution. In his paper, "Total Resource Budget Planning for Academic Libraries," Sherm Hayes explores not only the "operating" and "cash" budgets but also the "remainder" budget, consisting of contributed services and goodwill.

The final group of papers discusses "New Opportunities" for academic and research libraries. These include rapidly changing technological developments, explored jointly by David Weber and Richard McCoy; cooperative and automated networks, discussed by JoAn Segal; external contracting of library services, promoted by Nina Cohen; grantsmanship and the attraction of donors, prescribed by Theodore F. Welch.

A bibliography of additional citations compiled by Nancy E. Elkington, short biographical sketches of the contributors, and a subject index conclude the book.

This is an unpretentious collection of information based on the experience, research, and reflections of academic administrators, librarians, and business executives. It does not give all the answers to the difficult contemporary questions of how to cope with technological changes and limited monetary resources. However, it does identify many relevant issues, offer several practical directions to pursue, and stimulate thought. The authors' perspectives on the common issues in academic and library effective survival are varied and provide a thoughtful analysis of the topic. The only disappointment on finishing the book is the realization that it cannot fully replace the informal exchange that must have occurred as part of the conference interactions. Recommended as a valuable souvenir of the meeting, as a practical gift for academic library directors' administrators, and as professional reading for those of us who stayed home.—*Danuta A. Nitecki, University of Maryland at College Park.*

Other Recent Receipts

Listed here are books and other publications received for review that are of potential interest to LITA members. Some of these materials may be reviewed in later issues of ITAL.

Crawford, Walt. *MARC for Library Use: Understanding the USMARC Formats.* Professional Librarian Series. White Plains, N.Y. and London: Knowledge Industry, 1984. 222p. ISBN: 0-86729-120-6, hardcover, $36.50; 0-86729-119-2, softcover, $28.50.

Essential Guide to the Library IBM PC. V.1, Nancy Jean Melin, *The Hardware: Set-Up and Expansion.* V. 2, Suzana Lisanti, *The Operating System: PC-DOS.* Westport, Conn.: Meckler, 1985–86. ISBN: 0-88736-033-5, spiralbound, $19.95 (V.1); 0-88736-034-3, spiralbound, $19.95 (V.2).

Heilprin, Laurence B. *Toward Foundations of Information Science.* White Plains, N.Y.: Knowledge Industry, 1985. 231p. ISBN: 0-86729-149-4, hardcover, $34.95. "Published for the American Society for Information Science."

Katz, Bill, ed. *Reference and Online Services Handbook: Guidelines, Policies, and Procedures for Libraries. Vol. II.* New York: Neal-Schuman, 1986. 602p. ISBN: 0-918212-49-9, hardcover, $39.95.

Keller, Dean H., ed. *Libraries in the '80s: Papers in Honor of the Late Neal L. Edgar.* New York and London: Haworth, 1985. 157p. ISBN: 0-86656-459-4, hardcover, $29.95. "Also published as *Technical Services Quarterly,* Volume 3, Numbers 1/2."

Lynch, Beverly, ed. *Management Strategies for Libraries: A Basic Reader.* New York and London: Neal-Schuman, 1985. 682p. ISBN: 0-918212-86-3, softcover, $35.

Marcus, Marvin, and Rebecca Marcus. *MacAlgebra: BASIC Algebra on the Macintosh.* Computers and Math Series. Rockville, Md.: Computer Science Pr., 1986. 429p. ISBN: 0-88175-135-9, spiralbound, $24.95; student's

diskette and teacher's diskette available for $20 each.

Murr, Lawrence E. and others. *Information Highways: Mapping Information Delivery Networks in the Pacific Northwest*. Portland, Oreg.: Hypermap, 1985. 77p. Softcover, $25.

Nolan, Jeanne M. *Micro Software Evaluations*. Lib. ed., V.2, 1985. Westport, Conn., and London: Meckler, 1986. 152p. ISBN: 0-88736-032-7, softcover, $95; ISSN: 8755-5794.

Rowley, J. E. *Computers for Libraries*. 2d ed. London: Bingley, 1985. 195p. ISBN: 0-85157-388-6, hardcover, $22.50 from Shoe String, Hamden, Conn.

Van Young, Sayre. *MicroSource: Where to Find Answers to Questions about Microcomputers*. Littleton, Colo.: Libraries Unlimited, 1986. 220p. ISBN: 0-87287-527-X, softcover, $22.50 in U.S., $27 elsewhere.

Wood, M. Sandra, ed. *Cost Analysis, Cost Recovery, Marketing, and Fee-Based Services: A Guide for the Health Sciences Librarian*. New York: Haworth, 1985. 268p. ISBN: 0-86656-353-9, hardcover, $29.95. "A monographic supplement to the journal *Medical Reference Services Quarterly*, Volume 4, 1985." ■■

Your invitation to enter the
1987
JOHN COTTON DANA LIBRARY PUBLIC RELATIONS AWARDS CONTEST

IF YOU'VE DONE an outstanding job of making your community more aware of your library, the John Cotton Dana Library Public Relations Awards Contest can tell the world about your efforts.

Your entry will be considered among those from libraries of all types, sizes, and budgets. Entries are judged by a panel of your peers, and two types of awards are given.

The John Cotton Dana Award
This award is given for a library's total annual, coordinated public relations program, including publicity, programs, advertising, publications, exhibits, special events, promotions, and audio-visual presentations.

The Special Award
The Special Award is given in recognition of a part of your public relations program—a fund-raising campaign, a series of adult or children's programs, or any other special project.

Contest Dates
Entries for the 1987 John Cotton Dana Library Public Relations Awards Contest may reflect any one of the following time frames:

- Calendar year 1986 (January-December)
- School Year 1985/6 (Fall-Spring)
- Special project which ends in 1986.

The deadline for entries is February 2, 1987.

Awards Ceremony
Official award citations will be presented to contest winners at the 1987 annual conference of the American Library Association, at a reception hosted by The H.W. Wilson Company.

Sponsorship
The John Cotton Dana Library Public Relations Awards Contest is sponsored jointly by The H.W. Wilson Company and the Public Relations Section of the Library Administration and Management Association, a division of the American Library Association.

To Enter
To request an Information Packet containing contest entry forms, rules and regulations, questions and answers about the awards, a sample of the judges' evaluation form, names of the contest judges, and a list of previous winners, please write to: Library Relations Department, The H.W. Wilson Company, 950 University Avenue, Bronx, NY 10452.

CATALOGING?

HARRIET VELAZQUEZ
Executive Vice President

LET ME TELL YOU WHY UTLAS HAS THE BEST ONLINE CATALOGING SUPPORT SYSTEM.

1. OWNERSHIP
You own your records. Utlas' client libraries maintain their own files so they can edit records as they choose and add local holding codes and other information as required.

2. INTERNATIONAL DATABASE
Utlas has over 24 million MARC records and over 5 million REMARC records including records from the Library of Congress, British Library, National Library of Canada, Government Printing Office, National Library of Medicine, Utlas customers and many other sources.

3. FREEDOM OF CHOICE
The Utlas database retains the unique contributions of its client libraries. You choose the record to suit your need: from medical libraries with MESH headings to public libraries with modified LC headings.

4. AUTHORITY CONTROL
Utlas has in fact what so many others claim, real automated authority control. Consistent AACR2 headings and linked authority records for cross references. LC names, subjects and series are up to date.

5. CUSTOM-TAILORED PRODUCTS
Utlas can create catalog cards, magnetic tapes and catalogs from your file. Products are available in laser-printed typeset or microform format.

UTLAS
EFFECTIVE SOLUTIONS FOR LIBRARIES WORLDWIDE

We can demonstrate why you should be using Utlas' CATalog Support System. CATSS is the leading edge. Call us today.

Utlas International U.S. Inc., 1611 North Kent Street, Suite 910, Arlington, VA. 22209 Telephone (703) 525-5940 or (800) 368-3008