

LIBRARY RESOURCES & TECHNICAL SERVICES

VOLUME 36, NUMBER 2

APRIL 1992

ARTICLES

- | | | |
|-------------------------------|-----|----------------------------------------------------------------------------------------------------------------------------|
| Gillian M. McCombs | 135 | Technical Services in the 1990s: A Process of Convergent Evolution |
| Josefa Abrera and Debora Shaw | 149 | Frequency of Use of Cataloging Rules in a Practice Collection |
| Barbara B. Tillett | 162 | Bibliographic Relationships: An Empirical Study of the LC Machine-Readable Records |
| Gregory H. Leazer | 189 | An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Schema for the USMARC Formats |
| Edgar A. Jones | 209 | Consistency in Choice and Form of Main Entry: A Comparison of Library of Congress and British Library Monograph Cataloging |
| Arlene G. Taylor | 224 | Variations in Personal Name Access Points in OCLC Bibliographic Records |

FEATURES

- | | | |
|-----------------------------|-----|------------------------------------------------------------|
| | 242 | <i>LRTS</i> 1991 Referees |
| Lawrence W. S. Auld, Editor | 243 | Book Reviews |
| Alan R. Thomas | 243 | <i>The Bibliographic Record and Information Technology</i> |
| Margaret E. Doutt | 244 | <i>Describing Archival Materials</i> |
| Fred W. Jenkins | 244 | <i>Early Bindings in Paper</i> |
| Don Lanier | 245 | <i>Legal and Ethical Issues in Acquisitions</i> |
| Diane D. Kester | 246 | <i>Library Cooperation and Networks</i> |
| Constance Rinehart | 247 | <i>Library of Congress Subject Headings</i> |
| Ellen Crosby | 248 | <i>Library Technical Services</i> |
| Lori Osmus | 250 | <i>Manheimer's Cataloging and Classification</i> |
| | 252 | Letter |
| | 253 | Instructions for Authors |
| | 252 | Index to Advertisers |

36(2):129-256

ISSN 0024-2527

AMERICAN LIBRARY ASSOCIATION

ASSOCIATION FOR LIBRARY COLLECTIONS & TECHNICAL SERVICES

EDITORIAL BOARD

Editor and Chair of the Editorial Board RICHARD P. SMIRAGLIA
Editorial Assistant GREGORY H. LEAZER
Assistant Editors:

KAREN A. SCHMIDT for Acquisition of Library Materials Section
CHARLES SIMPSON for Cataloging and Classification Section
MICHAEL T. RYAN for Collection Management and Development Section
CARLA J. MONTORI for Preservation of Library Materials Section
THOMAS A. BOURKE for Reproduction of Library Materials Section
MIRIAM PALM for Serials Section
D. KATHRYN WEINTRAUB Special Editor
EDWARD SWANSON Special Editor
LAWRENCE W. S. AULD Book Review Editor

Ex-Officio Members:

JOAN W. HAYES, Chair, Council of Regional Groups
KAREN MULLER, Executive Director, ALCTS

Library Resources & Technical Services (ISSN 0024-2527), the quarterly official publication of the Association for Library Collections & Technical Services, a division of the American Library Association, is published at ALA Headquarters, 50 E. Huron St., Chicago, IL 60611. *Business Manager:* Karen Muller, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Send manuscripts to the *Editorial Office:* Richard P. Smiraglia, editor, *Library Resources & Technical Services*, School of Library Service, Columbia University, 516 Butler Library, New York, NY 10027 (BITNET User ID: smiragli@cunivm). Advertising Sales Manager, Stuart M. Foster; Advertising Coordinator, Connie Barone, c/o *Choice*, 100 Riverview Center, Middletown, CT 06457, phone (203) 347-6933. *ALA Publishing Services:* David Epstein, Eileen Mahoney, Dianne M. Rooney; Production: Donovan Vicha, Amy Brown, Bruce Frausto, Josephine Gibson-Porter, and Daniel Lewis. *Subscription Price:* to members of the Association for Library Collections & Technical Services, \$22.50 per year, included in the membership dues; to nonmembers, \$45 per year in U.S., \$55 per year in Canada and other foreign countries. Single copies, \$14.

Second-class postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611.

Library Resources & Technical Services is indexed in *Library Literature*, *Library & Information Science Abstracts*, *Current Index to Journals in Education*, *Science Citation Index*, and *Hospital Literature Index*. Contents are listed in *CALL* (*Current American—Library Literature*). Its reviews are included in *Book Review Digest*, *Book Review Index*, and *Review of Reviews*.

Instructions for authors appear on p. 253-54 of the April 1992 issue of *Library Resources & Technical Services*. Copies of books for review should be addressed to book review editor, Lawrence W. S. Auld, Department of Library and Information Studies, 215 Joyner Library, East Carolina University, Greenville, NC 27858-4353.

© American Library Association 1992

All materials in this journal subject to copyright by the American Library Association may be photocopied for the noncommercial purpose of scientific or educational advancement granted by Sections 107 and 108 of the Copyright Revision Act of 1976. For other reprinting, photocopying, or translating, address requests to the ALA Office of Rights and Permissions, 50 E. Huron St., Chicago, IL 60611.

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

Publication in *Library Resources & Technical Services* does not imply official endorsement by the Association for Library Collections & Technical Services nor by ALA, and the assumption of editorial responsibility is not to be construed as endorsement of the opinions expressed by the editor or individual contributors.

SUMMA CUM LAUDE . . .

is how academic librarians rank EBSCO's serials management services! That's because we've helped the staffs of hundreds of academic and research libraries reach the highest level of excellence and efficiency in serials acquisition, management and control. Our *Academic & Research Library brochure* tells the whole story. Contact your EBSCO representative or the Regional Office nearest you to receive your free copy. Or write us at the address below.



International Headquarters

P.O. Box 1943

Birmingham, AL 35201-1943

(205) 991-6600 • FAX (205) 991-1479

The Wilson Indexes on
Magnetic Tape!

WILSON TAPE[®]

DATABASE LICENSING SERVICE

Versatility-Accessibility-Affordability

Bring the renowned Wilson Indexes to your institution with easy, unlimited local electronic access. *WILSON TAPE* provides machine-readable monthly tapes for each database—in USMARC or Wilson-tagged format—making the data accessible with the same system of commands and protocols as your online catalog.

The Premier Choice for System-Wide Data Access

WILSON TAPE Service provides access to Wilson's renowned databases on a 24-hour basis. Patrons can access data from homes, offices, or dormitories with the same hardware and software used for the library's public access catalog or other locally-loaded databases. The number of simultaneous users is limited only by system capacity and you pay the same annual fee regardless of how frequently the files are accessed.

Create Your Own Customized Electronic Library

By selecting the databases that best meet your needs, you create a customized databank that:

- Maximizes the use of existing hardware and software
- Enhances use of periodical and book holdings
- Promotes simultaneous use by multiple searchers on a round-the-clock basis
- Enables unlimited off-site access to centralized reference holdings
- Reduces training time by providing easy access to data with familiar search software
- Eliminates costly connect-time charges

Customized Databanks at Individualized Rates

With an innovative, easy-to-understand pricing structure, WILSON TAPE Service is affordable to institutions of every size. An initial subscription to any database tape includes data from 1989, plus monthly updates for the forthcoming year. Retrospective data prior to 1989 is also available.

**Call Toll-Free
1-800-367-6770, Ext. 2758,**
for more information and
a price quotation.

Telefax
1-212-590-1617.



THE H. W. WILSON COMPANY

950 UNIVERSITY AVENUE, BRONX, NY 10452

Technical Services in the 1990s: A Process of Convergent Evolution

Gillian M. McCombs

The theory of convergent evolution in the life sciences is used as a metaphor to illustrate the heightened levels of integration achieved by technical services, public services, and collection development in academic and research libraries. The vehicle for this development is automation, which has evolved to the stage where migrations from one automated system to another are more common than new system implementations, and a complete rethinking of information processing is emerging. Organizational charts have been slow to reflect these changes, but workflow patterns are showing more reliance on matrix organizational theories than on traditional hierarchies. It is difficult to see clearly the direction technical services librarians should be moving and to plot the steps necessary to move in that direction. What is clear, however, is that new information needs and changing library environments are revitalizing and restructuring traditional internal relationships.

In the field of life sciences, the term *convergent evolution* is used to describe the process whereby two unrelated and dissimilar species develop similar attributes so that they come to resemble each other. This development results from "similarities in the habits of the organism or in the environment."¹ In anthropology, the same term is used to denote the "independent development of similarities between unrelated cultures."² This natural phenomenon is a useful metaphor for illustrating the changing nature of the relationship between technical and public services over the last fifteen years.³ Public services and technical services, recently joined by col-

lection development, are actually beginning to function together as one, evidencing the multiplicity of convergent developments and at long last producing the "ecumenical librarian" propounded by Michael Gorman in 1983.⁴

BACKGROUND

The long-established monophyletic approach of the profession was described by Gorman in 1979, when he humorously referred to the two species as "the sheep and the goats."⁵ The gradual move toward a polyphyletic approach, as some academic librarians have become less specialized and more multifunctional, has been

GILLIAN M. MCCOMBS is Assistant Director, Technical Services, University at Albany, State University of New York. The author wishes gratefully to acknowledge Peter Vonnegut, Thomas E. Dewey Graduate Library for Public Affairs and Policy, University at Albany, State University of New York, who assisted in the preparation of the diagrams. Manuscript received February 8, 1991; revised October 15, 1991; accepted for publication December 16, 1991.

documented voluminously.⁶ For several years, progress was hampered by a variety of factors, such as the need for technical services and public services to inhabit different work areas,⁷ the turf-oriented nature of middle management,⁸ the general resistance to change evidenced in the profession as a whole,⁹ and the barriers imposed by the different automated systems used in each department of the library.¹⁰ There were expectations with the early automated library systems (such as LCS, CLSI, Geac, and LS 2000) that workflow was going to change. However, we have learned by looking back over these early automation efforts that there are a number of evolutionary stages that occur before workflow really changes. As De Klerk and Euster document,

Typical stages in technological innovation include an initial period during which the individual manual processes, that may or may not be combined, are emulated. During this first experimental stage, that mechanized or automated mode is "layered" on to the manual process and both are in use. In a later stage, the manual process will be abandoned. Automated processes then substitute for manual ones but typically in the same context as before. The total system will not be completely rethought for some time, and advantage is not taken of new possibilities for fresh sequences and combinations.¹¹

Early compilations of organization charts show the traditional divisions of public and technical services and the insular nature of most librarians' work.¹² In 1985, the Association of Research Libraries (ARL) conducted a survey to look at library organization specifically in the context of automation.¹³ The results of the survey indicated that although the traditional patterns had changed little, there continued to be expectations that major changes were imminent; there was much discussion but still little actual integration. The assistant directors of public and technical services divisions were looking to maximize the opportunities afforded by automation, both in managerial areas (staff savings) and service areas (capabilities for new and improved services). Another sur-

vey, this time by De Klerk and Euster in 1989, looked at library organization from the top down. Fifty-three library directors were asked informally for their perceptions on automation as a catalyst for organizational change. The conclusions were that there were a number of changes being made but that no single pattern could be detected: "The present spectrum of changes in library organizations strongly points to today as a period of experimentation—one in which a variety of forms are being tried in an effort to increase coordination and flexibility."¹⁴ The very next year, Patricia Larsen surveyed more than two hundred university libraries, asking how many had actually changed their organizational structures and, if so, why. The conclusions, to be published in *The Reference Librarian*, number 34, were again that although "the long standing divisional structure is still very much an accepted, viable organizational pattern," an increasing number of institutions had made substantial changes or were in the process of doing so, and "integration of functions was apparent at several points in the survey responses."¹⁵ Documented crossover between public and technical services librarians was increasing, leading one to believe that convergent evolution might well be occurring. Sixty-seven percent of the respondents indicated that they believed their libraries had become more integrated since automation; the author illustrates this with an accompanying chart that shows the areas and tasks where this is happening. The most recently published ARL publication on organization charts (1991) corroborates these conclusions: "Internally, library organizations, traditionally departmentalized by function, continued to experiment with a matrix organizational structure."¹⁶ The presumption is that this integration is achieved by staff willing to cross over or break down traditional barriers, move into new areas, and learn unfamiliar responsibilities.

It is also interesting to note that one of the survey results reported in the subsequent ARL Systems and Procedures Exchange Center (SPEC) Kit *Training of Technical Services Staff in the Automated*

Environment showed that "41.4% of the respondents had replaced the first system and now were working with the second or even the third or fourth system."¹⁷ It seems that a new phase in automation has been reached, thereby providing the vehicle for convergent developments. As libraries purchase their second or third systems, and "migrate" rather than "implement" becomes the watchword, a move toward both the concept of the "renaissance librarian"¹⁸ (which would presume a certain amount of convergence) and the truly integrated library information system seems visible. Catalogers still "go forth and classify," but they no longer "dwell in the darkness"; their ways are no longer "secret and hidden."¹⁹ Although the amount of time spent on functions other than cataloging is still fairly minimal and the traditional public and technical services splits are still the norm,²⁰ the number of academic libraries where catalogers are performing some public service on a regular basis, or have responsibilities in collection development, is growing.²¹ It is not clear, though, whether this is a result of a genuine convergent evolutionary growth, in which librarians themselves are changing, or whether it is the result of enforced redeployment and streamlining due to current budget cuts and staff shortages. If we are able to move to the next evolutionary stage in De Klerk and Euster's description of the process of technological innovation, and if the evolutionary process is genuine and not artificial, what will the revised organization charts look like?

THE TRADITIONAL ROLE OF TECHNICAL SERVICES

Technical services staff in an academic library catalog, process, and provide access to the materials housed in the library. When this definition is put in the context of the mission statements of most academic research libraries, which generally include a phrase such as "play an essential role in information access, analysis, organization, distribution and management," the centrality of technical services to the mission of any university library cannot be disputed.

What is done, or not done, in technical services has enormous, far-reaching impact. As Tauber states,

most practising librarians recognize that an effective organization of technical services is essential if the library is to provide its users with high quality services. . . . It has been well demonstrated that an effective technical services program is woven deeply into the fabric of efficient library service.²²

The major evolutionary changes that are taking place, however, are redefining the role technical services librarians will play in working toward these goals, as some of the functions that were previously the sole domain of centralized technical services are being done elsewhere in the library by other librarians as well as by paraprofessionals. For example, branch library staff are able to check in their own subscriptions; bibliographers or collection development support staff can search a cataloging utility in the preorder process and download a record that will subsequently be used by cataloging staff; public services staff can access the system to answer questions that used to be phoned down to technical services for shelflist or Kardex information; reference librarians are developing software packages that link citation databases to local periodical holdings; and public services librarians use online reports of unsuccessful online searches in order to suggest additional access points.

Enabling this continued convergence of functionality is a technological framework that encompasses both the automation of technical services processing and administrative functions, and the automation of the traditional information-seeking and information-providing mechanisms in user services. All of these functions are now coming together in a truly integrated online library information system and providing the environment necessary for technical and public services librarians to move closer together. There are two levels at which these evolutionary changes are taking place: the first is the nitty-gritty workflow level, and the second is the campuswide level, i.e., the role to be played by

the library in the provision of information services to the university community.

WORKFLOW CHANGES

The traditional processing workflow in a large, semi-automated academic library consists of material being ordered, received, cataloged, and prepared for the stacks, each process or function being performed in a different unit within the technical services division (see figure 1). With a truly integrated system, however, these functions do not necessarily need to be performed by different departments or personnel (see figure 2). The item can be searched first during the collection development process (which in some institutions occurs in technical services and in

others in a separate collection development department) and the record found subsequently used for ordering and invoicing, for receipt and claiming, and then for cataloging. Work can be done by collection development, acquisitions, or copy-cataloging staff. Workflow in many large research libraries is already changing in this direction as increasingly integrated systems are installed.²³ Because on-order information, which was traditionally considered within the purview of technical services, is now available in the online catalog, the sense of "information ownership" that once existed is disappearing.

AUTOMATION CHANGES

The online systems supporting these two

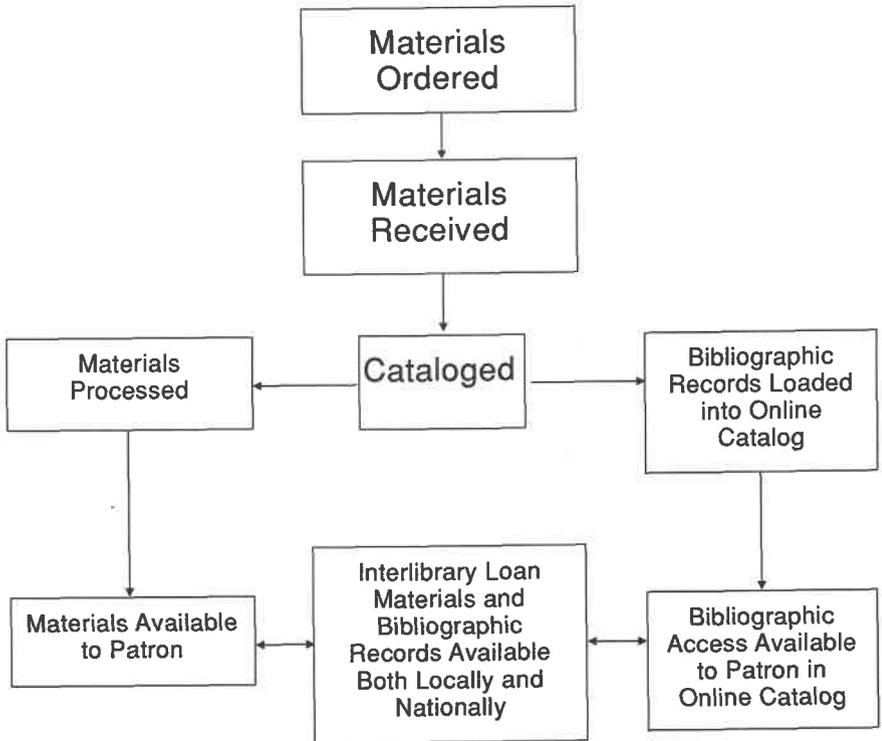


Figure 1. Technical Services Workflow (current).

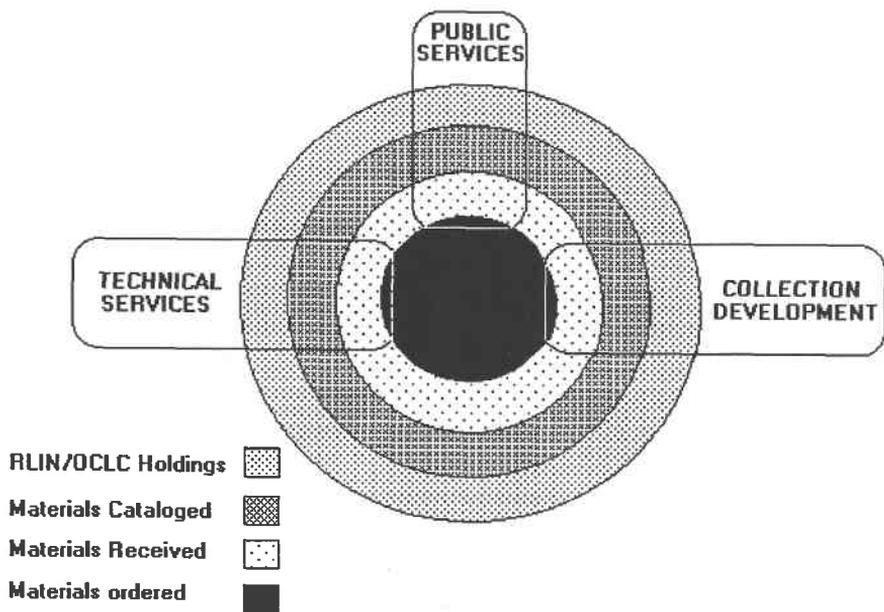


Figure 2. Technical Services Workflow (future).

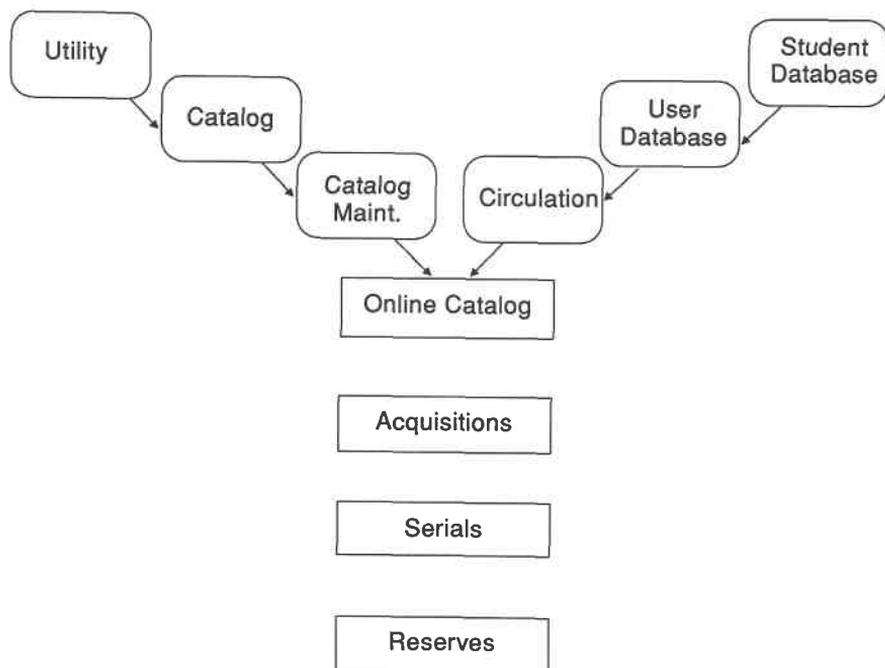


Figure 3. Library Information System (current).

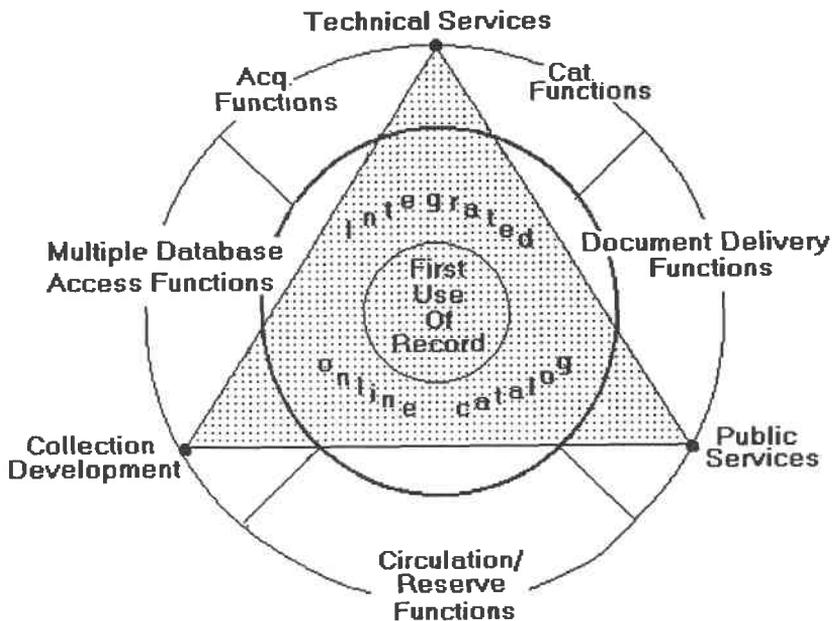


Figure 4. Library Information Services (future).

versions of workflow are illustrated in figures 3 and 4. Figure 3, representing current library systems, shows cataloging being done on the utility, with the bibliographic records being loaded into the cataloging subsystem, passed through catalog maintenance, and entered into the online catalog. A similar flowthrough happens with circulation information. The student database feeds into the user database and matches item and charge information with the bibliographic record. Acquisitions, serials control, and reserve functions have previously been accommodated in stand-alone systems not connected to the online catalog and often only accessed by a separate terminal or by exiting the online catalog and entering a separate subsystem.

The new, integrated library information system (see figure 4) resembles more an onion than a series of connected boxes and is relational rather than directional. The first use of a bibliographic record is when

it is downloaded from a cataloging utility or other source (see center of figure 4). The record is immediately available in the online catalog, represented by the next layer of access. The outer band in the diagram is where the other functions reside, all still accessing that one bibliographic record. With collection development involved in original selection of the cataloging record and in eventual use of that record at the reference desk (because many bibliographers also perform reference duties), some of the historic gaps between traditional public and technical services concepts will narrow. The system is providing the opportunity for convergent evolution to flourish. Collection development staff will move closer to catalogers as they are trained in cataloging standards and optimal record selection. Acquisitions staff will no longer have to input separate records for ordering purposes and could even "catalog" the item on receipt. Authority

control can be applied at a number of different stages in this process and will ensure consistency of access. The shaded triangle in figure 4 represents the joint functional use of the system by collection development, user services, and technical services. This tripartite relationship may well replace the traditional public and technical services divisions as all three functionalities move closer together.

Jasper Schad calls this theme of convergent evolution "a form of organizational Darwinism" in which multivariate assignments for collection development librarians produce something similar. He also evinces concern that this could be due more to heavy work loads than to a deliberate and genuine evolutionary trend.²⁴ There is in addition a large "turf" issue that may well need to be defined differently before it can be resolved satisfactorily. The "where" of cataloging an item is not a simple issue; neither is determining the level of staff needed to perform this task. Management needs to take a leadership role in objectively looking at system capabilities, analyzing workflow, and developing new information processing models.

CAMPUSWIDE CHANGES

These evolutionary developments will have an impact on the campus as a whole

through the increasing connectivity of the online catalog, or representation of the library's holdings, within the array of information services provided to the users. Very basic representations of current and future information services are shown in figures 5 and 6. In figure 5, which depicts current information services, the user goes to the traditional reference desk and is funneled either to the card or online public access catalogs (OPAC) for library holdings or to a variety of reference tools, including print materials and computer search services for information that may be located outside the library. Interlibrary loan (ILL) is the communications network that brings this externally located information to the user in the library. In figure 6, a representation of future information services (in fact now present in some libraries), a user services station directs users to the library information system. This is a one-stop information gateway providing automated access to a full gamut of databases using a common command language and with a communications network and document delivery system that uses telefacsimile and optical scanning as well as traditional ILL methods. The new *X-Window System*, providing high-performance, high-level, device-independent graphics, is already in use at several institutions. A hierarchy of resizable, overlapping windows allows the user to have on one screen records from

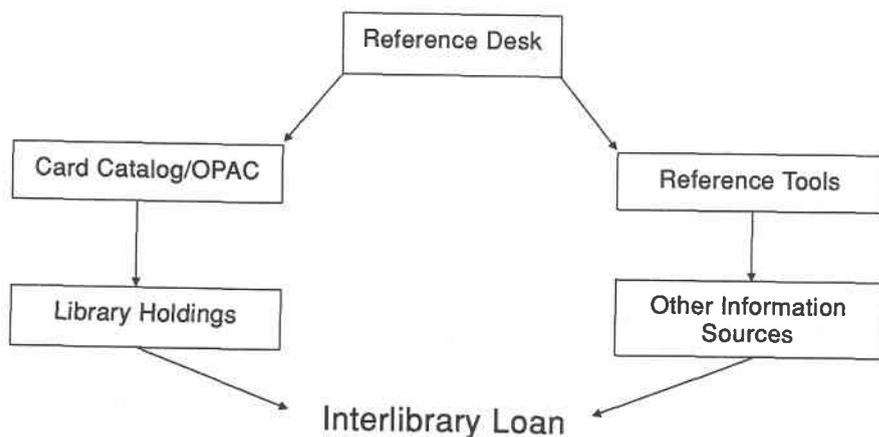


Figure 5. Library Information Services (current).

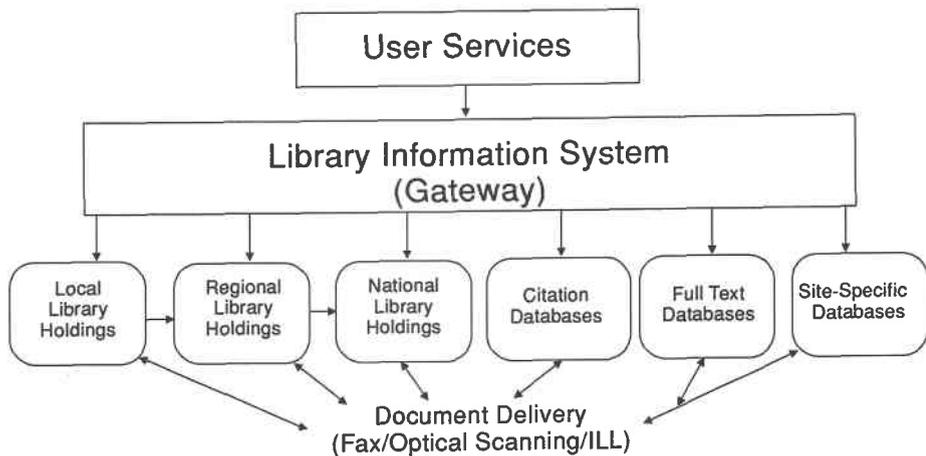


Figure 6. Library Information Services (future).

local, regional, and national library systems. This network-transparent access provides for functional separation with little degradation of response time, a crucial requirement for the distributed environment.²⁵

One of the major differences between figures 5 and 6 of prime significance for the user is the proliferation of databases other than local holdings. In figure 5, the library's holdings and its access points represent at least 50 percent of the figure, more when it is acknowledged that many of the print reference tools are library holdings and as such are accessed through the local catalog. In figure 6, the library's holdings database represents only one of the six different databases and information sources available. It is important to remember, however, that the library's holdings probably satisfy about 90 percent of current information requests, especially as many of the citation databases will refer patrons to material held in the library. Some systems provide "hooks" to indicate whether or not the library holds these journals. The capability also exists in many CD-ROM packages for manually adding a link to library-specific holdings. But as information-seeking behavior changes, as systems improve to allow off-campus access to all these services and to provide document delivery options that bypass the library (as well as the ability to pay

for these services off-site), the traditional concept of a library and its resources will indubitably change. The idea of acquiring on demand rather than in advance of use is much discussed in articles on the library of the future.²⁶ Recent statistics from the OCLC Online Computer Library Center, Inc., show that from 1987 through 1991, cataloging activity decreased 9.23 percent, while interlibrary loan requests increased 30.52 percent, a result of increases in both resource sharing and cataloging on local systems.²⁷ OCLC is also working on two research projects specifically in the area of document delivery—Project ADAPT and the Document Imaging Processing Toolbox. Cataloging in the Internet environment will require the rethinking of some basic tenets that will underscore the need for continued encouragement and support of convergence or role melding. In a series of interviews with the members of the OCLC Cataloging and Database Services Advisory Committee, opinions voiced ranged from predictions that cataloging as we know it will become unnecessary to the simple belief that the roles of original and copy catalogers are starting to blur.²⁸ Public services librarians are also concerned about their future, since both the Research Libraries Information Network (RLIN) and OCLC have developed pilot projects aimed specifically at end-user searching

for faculty—the Research Access Project (RLIN) and FirstSearch (OCLC)—that could result in the library and reference librarians being bypassed.

INTEGRATED FUNCTIONALITY

As collection development, user services, and technical services functions begin to converge, thus forming the triangle of common functionality illustrated in figure 6, organization charts will undoubtedly begin to change. Figure 7 shows a very simple organization chart using a matrix instead of the traditional hierarchical structure. The integration of all functions in one automated system allows for support staff to be pooled. The reduction in the number of different operating systems to be learned allows flexibility in training staff and stationing them wherever backlogs develop. Departments are looking to private industry for new models of problem solving in which, instead of all staff members in a department answering all levels of questions, a layered approach is employed, with a "bumping up" of the thorniest problems. Union rules may well

determine how these changes take place. An institution with more than one bargaining group might find it hard to replace support staff with paraprofessionals from another bargaining group. It might be that changes can be made as lines are lost or turnover occurs, facilitating the rewriting of job descriptions. This concept will result in a broader understanding of more parts of the puzzle by more of the players and is discussed both generally as an organizational alternative for libraries, as well as specifically, in the area of collection development.^{29,30} The University of Alberta Library already shows a matrix arrangement for its official organization chart, the overlapping ovals indicating areas of functional convergence (figure 8).³¹

CONVERGENCE

The level at which librarians are working together to satisfy users' needs is an indication of the extent to which roles are converging. It can sometimes happen in technical services, where there is little direct contact with the public, that user information needs can seem less important than

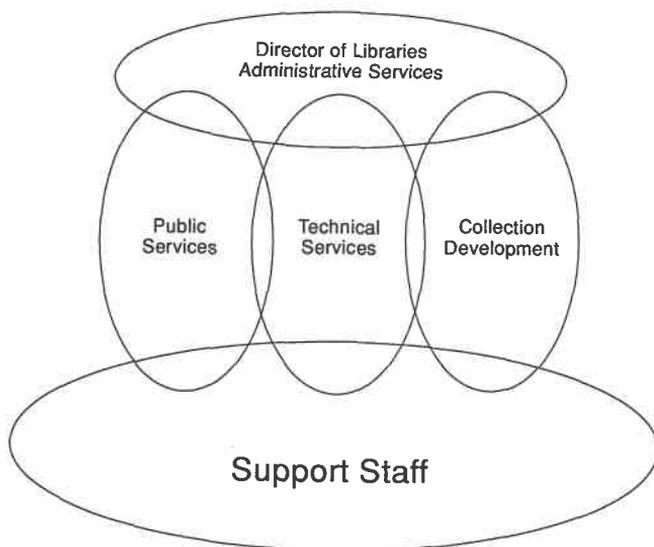


Figure 7. Matrix Staffing Chart Illustrating Areas of Convergence.

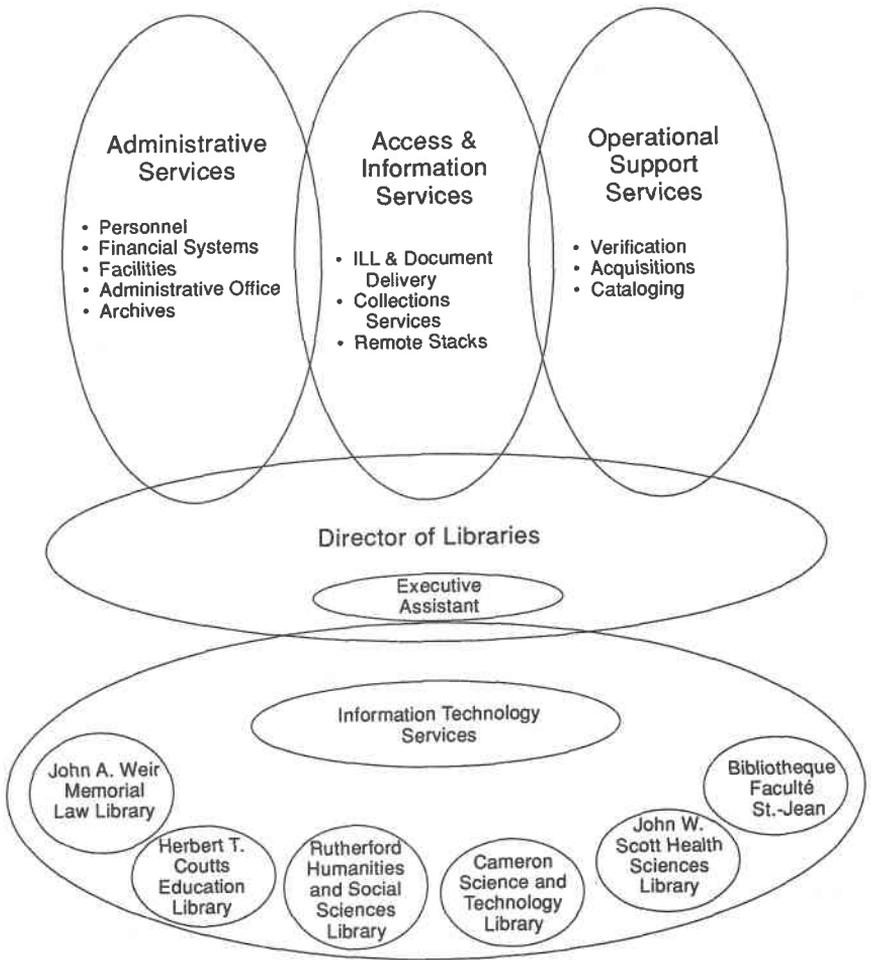


Figure 8. University of Alberta Library: Organizational Matrix (October 1990).

Library of Congress rule interpretations and the *Anglo-American Cataloguing Rules*. Interaction with the user is an essential ingredient for successful information processing. In the *Library Resources & Technical Services* review of descriptive cataloging research in 1990, Jay Lambrecht succinctly described the primary objectives of research into descriptive cataloging as being "to understand the needs of the users we serve and to discover and document the best means of meeting those

needs. By doing so, we retain control of our own discipline and move it forward."³²

The traditional technical services support activities have moved onto the reference desk in the shape of immediate online access to what used to be considered technical services files. If, as Lambrecht suggests, we continue to focus on improving process but conduct no research to understand the needs of users, then "we do not have a clear view of the larger picture. We cannot assume that we know what users

need from description in the online catalog, and we are not asking them."³³ By joining forces with our counterparts at the reference desk and learning their "secret ways,"³⁴ we will gain firsthand understanding of what users' needs are. One method that libraries are actively employing to facilitate convergence and enable all types of librarians to work together is the implementation of some form of reference intern program. Staff members from different parts of the library, both librarians and paraprofessionals, are trained to fill in at the reference desk during regularly scheduled reference department meetings. Another method is the creation of a catalog information desk using, among others, technical services librarians—although administrators are often averse to creating new service points in these days of shrinking staff resources. Catalogers learn to retrieve information from the online catalog, which they might have used sparingly before, and rethink how they create access points. Frequent questions for information consistently formatted in non-*Library of Congress Subject Heading (LCSH)* language might lead to the adding of local subject headings to records. This experience will also increase sensitivity to areas of research for catalogers, as suggested by Lambrecht. Organizational changes that are based on subject rather than function and that provide for this kind of interaction are in place at a variety of institutions, the University of Illinois at Urbana-Champaign and Penn State being the most widely known.

REASSESSMENT OF THE STATUS QUO

In order for librarianship to evolve comfortably, several issues need to be resolved as functionalities merge, such as how to manage the library in times of fiscal austerity (such as the present), when the new services generate increased staffing needs, and how and whether to maintain the standards we have upheld for so long if tasks are being performed by staff who are not as specialized and have multiple roles. In some libraries, basic technical services housekeeping procedures, such as claiming journals, searching for out-of-print

items, monitoring the duplicate rate, and managing gifts and exchange programs, are going by the wayside as remaining staff resources are carefully meted out in order to "get the material on the shelves." Questions are being asked about whether or not we should continue to do many of the things we have been doing for the last twenty years. As Peter Graham said in his article "Quality in Cataloging: Making Distinctions,"³⁵ we all know that "quality in cataloging is inversely proportional to cataloging productivity," and we need to take some long and hard looks at the needs of the patron and the technology we have at our command and put our daily work in this context. Discussions have been opened by Carol Mandel in the ALA Technical Services Directors of Large Research Libraries Discussion Group as to the need for a new philosophy of "cataloging for access" as opposed to "cataloging for collections."³⁶ The "cultural lag" described by American sociologist William Ogburn in 1922,³⁷ i.e., the slow adjustment of social and cultural routines to rapidly changing technology, could just as easily be applied to technical services librarians, who, through the cataloging codes, are using technology to process, but not radically change, bibliographic description. Mary Bolin, head of cataloging at the University of Idaho, focuses specifically on the attitudes of the individual cataloger as she addresses the perennial problem of high standards in a production-oriented setting:

In cataloging, as in many other activities, it is hard to keep the broad goal of our endeavor in mind at all times, and easy to slip into the rote and thoughtless application of rules.

While applauding current efforts to streamline and simplify cataloging on a national level, she favors a holistic approach which is more than

viewing [one's job] as multifunctional; it is seeing how each task fits in with the range of services provided in the library, and how each contributes to the library mission.

Balancing quality and quantity is a difficult task, but she believes that

only if we truly get nothing but esoteric materials that don't seem to fit any categories do we have an excuse for cataloging slowly.³⁸

At the same time we are being asked to do more—add new services, deal with new formats, learn new systems. We must look to the management literature in general for solutions that can help us maintain effectiveness during chilly fiscal times.³⁹ This tendency to focus on the task at hand as though it were an end in itself was criticized exactly fifty years ago in Andrew Osborn's historic "Crisis in Cataloging" paper⁴⁰ and was recently discussed again by H. M. Gallagher:

In the turning of a kaleidoscope elements become rearranged and take on a new design. Similarly, with "Crisis," catalogers shift their focus from doing their *tasks well*, to considering *how well* their tasks contribute to the ends and purposes of the library.⁴¹

Sir John Harvey-Jones cautions managers against cutting everything by 10 percent as resources dwindle in economically tight times. Instead, he favors determining core services, reinforcing them, and cutting everything else by 15 percent.⁴² Technical services divisions are looking at revised, streamlined workflow that utilizes students for routine activities and relieves regular staff for problem solving and more complex activities. Cross-training has been going on in many libraries for some time and more opportunities will be created.⁴³ Cross-functionality will develop naturally as staff are trained to work in more than one subsystem and in more than one department. This will not only provide variety but widen the understanding of what we do and why we do it. Flexibility to move individuals from one work team to another will enable us to redefine on-the-job relationships.⁴⁴ We must learn to develop a "synectic orientation"⁴⁵ that allows us to look at the familiar and see it in new ways—the "kaleidoscope" of H. M. Gallagher.⁴⁶ Similarly, public services librarians are redefining their roles, reassessing the level of staff required at the reference desk, learning to decipher the MARC formats, and taking computer programming courses that will better enable them to serve the user.

CONCLUSION

As librarians interact in the organizational culture and their roles begin to converge,

they must develop a tolerance for ambiguity. They must learn to accept "fuzzy roles" and understand that success in formulating new organizational structures and the roles they play therein is often measured in years rather than months. They must be positive in communicating this to support staff and help alleviate on-the-job frustrations, because this state of flux will cause some staff members to develop a state of extreme job insecurity and anxiety. Planning will increase in importance, being essential in order to avoid continual crisis management. Varying scenarios should be developed for staffing, budget, and the utilization of remaining resources. The equation seems to change on a daily basis, and librarians will have ample opportunity to be creative, innovative, and responsible. Although we must develop the ability to see the long-term value of a project and not be impeded by the urge to cling to the old ways of doing things, it is also our responsibility to assess the organizational culture, the degree of tolerance for innovation, our capacity for entropy, and our tolerance for conflict—in other words, the key dimensions of the work place that will determine the success or failure of the best planned project.⁴⁷

The convergence of functionalities will in turn give rise to a new professional philosophy of integration. We can see the day-to-day integration of job functions, the technological integration of the systems with which we work, and the organizational integration of traditional hierarchies. There is no question that the interests of our patrons and academic communities are best served by a series of internal relationships that neither polarize nor politicize the questions of access; a harmonizing relationship in which there are only systems navigators who have discovered that the earth is not flat but round, that a working harmony where no one function is more important than the other is preferable to a divided world.

Cycles of convergence and divergence can be seen in most creative professions and are essential for separating the forces for change and stability in order to reconcile them better. Species that do not evolve in response to a changing environment do

not survive, dinosaurs being a case in point. Librarians have the opportunity to choose to evolve, to emphasize relationships, and to respond to the environment. Unless we pause to take stock, to reach out to librarians in other parts of the library, to sound out our users as to their informational needs, and to develop a philosophy more in keeping with those needs, we will find ourselves, like the dinosaurs, relics of the past instead of active participants in the information services of the future.

Scientists have pointed out that

The evolutionary arrow of time is a broken one—if we arrange all the available fossils in chronological order, they do not form a sequence of scarcely perceptible changes, like consecutive frames of a cine film, but instead contain seemingly discontinuous leaps.⁴⁸

Do we have the courage to attempt to define our own destiny and make one of those "discontinuous leaps"?

REFERENCES AND NOTES

1. *Dictionary of Life Sciences*, 2d ed., ed. E. A. Martin (London: MacMillan, 1983), p.87.
2. *McGraw Hill Dictionary of the Life Sciences*, ed. Daniel Lapedes (New York: McGraw Hill, 1976), p.198.
3. Denise P. Kaplan, "The Year's Work in Technical Services Automation, 1989," *Library Resources & Technical Services* 34:299-312 (1990). The ongoing merger of technical and public services functions was singled out as the most significant issue of the year.
4. Michael Gorman, "The Ecumenical Library," *The Reference Librarian* 9:55-64 (Fall/Winter 1983).
5. Michael Gorman, "On Doing Away with Technical Services Departments," *American Libraries* 10:435 (1979).
6. See bibliography in Kaplan, "The Year's Work in Technical Services Automation, 1989," and a series of articles reflecting diverse viewpoints on the topic "Should Reference Librarians Tamper with the Online Catalog?" *Journal of Academic Librarianship* 12:340-49 (1987).
7. James G. Neal, "The Evolving Public/Technical Services Relationship: New Opportunities for Staffing the Cataloging Function," in *Recruiting, Educating, and Training Cataloging Librarians: Solving the Problems*, ed. Sheila S. Intner and Janet Swan Hill (New York: Greenwood, 1989).
8. Jennifer Cargill, "Integrating Public and Technical Services Staff to Implement the New Mission of Libraries," *Journal of Library Administration* 10, no.4:21-31 (1989).
9. Neal, "Evolving Public/Technical Services Relationship," p.111-19.
10. *Ibid*, p.111.
11. Ann De Klerk and Joanne R. Euster, "Technology and Organizational Metamorphoses," *Library Trends* 37:457-68 (1989).
12. *Organization Charts*, SPEC Kit 1 (Washington D.C.: Association of Research Libraries, Office of Management Studies, 1977); *Organization Charts*, SPEC Kit 129 (Washington D.C.: Association of Research Libraries, Office of Management Studies, 1986).
13. *Automation and Reorganization of Technical and Public Services*, SPEC Kit 112 (Washington D.C.: Association of Research Libraries, Office of Management Studies, 1985).
14. De Klerk and Euster, "Technology and Organizational Metamorphoses," p.468.
15. Patricia M. Larsen, "The Climate of Change: Library Organizational Structures, 1985-90," *The Reference Librarian*, no.34 (in press).
16. *Organization Charts in ARL Libraries*, SPEC Kit 170 (Washington D.C.: Association of Research Libraries, Office of Management Studies, 1991).
17. *Training of Technical Services Staff in the Automated Environment*, SPEC Kit 171 (Washington D.C.: Association of Research Libraries, Office of Management Studies, 1991).
18. Gillian McCombs, "Public and Technical Services: Disappearing Barriers," *Wilson Library Bulletin* 61, no.3:25 (Nov. 1986).
19. Gorman, "On Doing Away with Technical Services," p.435.
20. Patricia A. Eskoz, "Catalog Librarians and Public Services—A Changing Role?" *Library Resources & Technical Services* 35:76-86 (1991).
21. Neal, "Evolving Public/Technical Services Relationship"; Cargill, "Integrating Public and Technical Services Staff."
22. Maurice F. Tauber and Associates, *Technical Services in Libraries: Acquisitions, Cataloging, Classification, Binding, Photographic Reproduction, and Circulating Operations* (New York: Columbia Univ. Pr., 1954).

23. Judith Niles, "Technical Services Reorganization for an Online Integrated Environment," *Cataloging & Classification Quarterly* 9, no.1:11-17 (1988).
24. Jasper G. Schad, "Managing Collection Development in University Libraries that Utilize Librarians with Dual-Responsibility Assignments," *Library Acquisitions: Practice & Theory* 14: 165-71 (1990).
25. Some libraries and vendors that are already making use of these developments are Carnegie Mellon University's Library Information System; the OhioLINK system in the Ohio State University system; NOTIS Paclink in New York State and Indiana; the RLG Ariel Project; and OCLC's FirstSearch.
26. Clifford A. Lynch, "Visions of Electronic Libraries," in *The Bowker Annual Library and Book Trade Almanac, 1990/1991* 36 (New York: Bowker, 1991), p.75-82.
27. David N. Forsythe, "Acquisitions or Access: A Changing Pattern?" *SOL Newsletter* 4, no.5/6: 2 (May/June 1991).
28. "Cataloging Experts Look to the Future," *OCLC Newsletter*, 192:22-23 (July/Aug. 1991).
29. Peggy Johnson, "Matrix Management: An Organizational Alternative for Libraries," *Journal of Academic Librarianship* 16:222-29 (1990).
30. Schad, "Managing Collection Development in University Libraries."
31. Figure taken from *Organization Charts in ARL Libraries*, SPEC Kit 170.
32. Jay H. Lambrecht, "Ours Should Be to Reason Why: Descriptive Cataloging Research in 1990," *Library Resources & Technical Services* 35:257-64 (1991).
33. Lambrecht, "Ours Should Be to Reason Why," p.262.
34. Gorman, "On Doing Away with Technical Services," p.435.
35. Peter S. Graham, "Quality in Cataloging: Making Distinctions," *Journal of Academic Librarianship* 16:213-18 (1990).
36. Carol Mandel, "Cataloging for Access," *The Reference Librarian*, no.34 (in press).
37. William Ogburn, *Social Change With Respect to Culture and Original Nature* (New York, B.W. Huebesch, 1922).
38. Mary K. Bolin, "Make a Quick Decision in (Almost) All Cases: Our Perennial Crisis in Cataloging," *Journal of Academic Librarianship* 16:357-61 (1991).
39. Jerry W. Askew and Beverly D. Anderson, "Maintaining Organizational Effectiveness During Periods of Financial Decline," *College Student Affairs Journal* 9, no. 2:14-21 (1989).
40. Andrew Osborn, "The Crisis in Cataloging," *The Library Quarterly* 11:393-411 (1941).
41. H. M. Gallagher, "Dr. Osborn's 1941 'The Crisis in Cataloging': A Shift in Thought toward American Pragmatism," *Cataloging & Classification Quarterly* 12, nos. 3/4:25 (1991) [emphasis original].
42. Sir John Harvey-Jones, *Getting It Together* (London: Heinemann, 1991).
43. Sheila Creth, "Staff Sharing—A Concept that Works: A Case Study" *Journal of Academic Librarianship* 10:287-89 (1984).
44. Gerald R. Lowell and Maureen Sullivan, "Self-Management in Technical Services: The Yale Experience," *Library Administration & Management* 4:20-23 (1990).
45. Joan Rapp, "Personnel Selection for Cataloging," *Library Resources & Technical Services* 34:95-99 (1990).
46. Gallagher, "Dr. Osborn's 1941 'The Crisis in Cataloging,'" p.25.
47. Rapp, "Personnel Selection for Cataloging."
48. Peter Coveney and Roger Highfield, *The Arrow of Time: A Voyage Through Science to Solve Time's Greatest Mystery* (New York: Fawcett Columbine, 1990), p.255.

Frequency of Use of Cataloging Rules in a Practice Collection

Josefa Abrera and Debora Shaw

A practice cataloging collection of 716 books was cataloged using Anglo-American Cataloguing Rules, second edition, 1988 revision, with each rule use recorded. A total of 20,247 rule uses was required, but of the 818 rules in the code, only 232 (28.4%) were used. Most frequently used were rules for choice of name (22.1A and 22.1B) and entry under surname (22.5A1). When rules are ranked by frequency of use, the distribution is best described by an exponential curve. When compared with other studies of rule use, the findings suggest that introductory cataloging instruction and expert systems can identify and focus on a core set of rules, safely ignoring those that are seldom used.

A cataloging code is important to librarians for at least three major reasons. First, from a managerial perspective, the code has a significant impact on how well and how expensively libraries provide access to resources. Recently Gregor and Mandel reviewed the challenges to librarianship from increasing demands on the catalog and limited resources for cataloging.¹ They emphasized the need to simplify cataloging and make it more cost-effective. A first step in any cost-benefit study is a description of the current environment: How is cataloging done? A second reason for studying the cataloging process is the need to understand practices in order to instruct new catalogers. While the official, codified exposition of the rules is provided by the *Anglo-American Cataloguing Rules*, second edition, 1988 revision (AACR2R), new catalogers can learn much by starting

with common applications and applying the cataloging code in familiar situations. A third reason for analyzing cataloging practices is to consider how this activity would be presented to an expert system. Meador and Wittig, among others, have examined how AACR2 can serve as a rule base for an automated cataloging system.² Hjerpe and Olander provided an insightful account and analysis of the structure that limits the use of AACR2 as a knowledge base.³

Underlying all of these approaches is an interest in discovering a "core set" of most frequently used rules. While humans can select applicable rules from the entire code, it helps to make explicit the most used or useful rules when preparing for automated applications such as expert systems. Meador and Wittig observed that "to date expert systems have not worked within a domain as large as the total AACR2

JOSEFA ABRERA and DEBORA SHAW are Associate Professors, School of Library and Information Science, Indiana University, Bloomington. The authors gratefully acknowledge the assistance of Terry Roy in data preparation and entry. Advice on statistical analysis provided by Charles H. Davis is also greatly appreciated. Manuscript submitted September 17, 1991; accepted for publication October 30, 1991.

rules would demand.⁴ One approach to trimming the cataloging code to fit into an expert system is to restrict the domain of expertise. Jeng looked at the expertise needed to determine the title proper, and others have looked at automatic approaches to determining access points.⁵⁻⁷ Davies noted that attempting to include the entire cataloging code in an expert system "would almost certainly be subject to the law of diminishing returns."⁸

The project described below focused on two questions. First, can a preliminary core set of rules for all aspects of descriptive cataloging be derived? Second, can further statistical analysis of rule use distributions assist in predicting the number and kinds of rules that should be emphasized in instruction or included in an expert cataloging system?

RULE USE DISTRIBUTIONS

In 1972 Fox reported on a study of the potential for automatic application of cataloging rules. She studied more than 12,000 Library of Congress proof slips, observing which rules from the 1949 *American Library Association Cataloging Rules for Author and Title Entries (ALA)* or the 1967 *Anglo-American Cataloging*

Rules (AACR) were used to state main entry.⁹ Richmond used Fox's findings, ranking the rules by frequency of use. She notes that the rank-frequency listing presents a hyperbolic distribution.¹⁰ Further analysis reveals that the curve of best fit for both sets of rules is a power curve ($y=ax^{-b}$), reflecting the great skew of the distribution caused by heavy use of the rule for personal author (see figures 1 and 2).

Fox's study and Richmond's analysis represent an important approach to investigating frequency of rule use through analysis of the cataloger's product. Fox's attention to automatic techniques anticipated current interest in the potential of expert systems for cataloging.

DESCRIPTION OF THE PROJECT

The entire practice cataloging collection at the Indiana University School of Library and Information Science was analyzed. The collection of 716 books was developed to provide a wide variety of cataloging exercises for students; for this reason the collection is probably more diverse than most general collections. However, 98% of the collection consists of English-language monographs. Thus, the findings might underestimate rule uses for nonbook materials

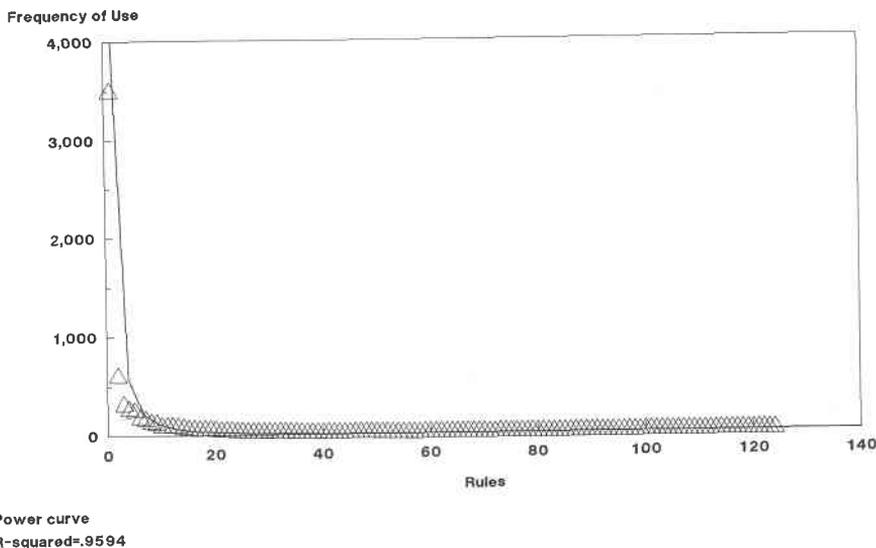


Figure 1. ALA Rules (Fox) Frequency of Use (n=124).

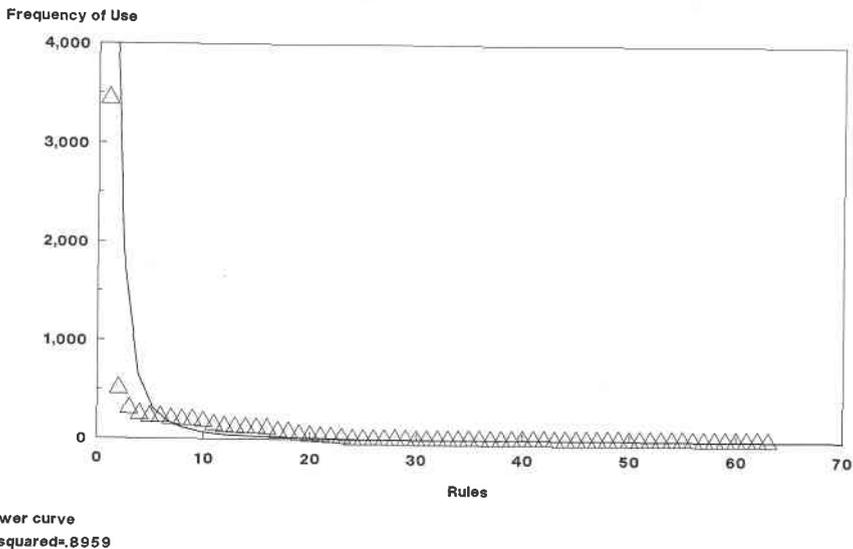


Figure 2. *Anglo-American Cataloging Rules (Fox) Frequency of Use (n=63).*

or collections with significant foreign-language holdings.

For each item in the collection, copy was sought in the database of the OCLC Online Computer Library Center, Inc., and its MARC record was printed. Patterns of authorship and type of main and added entries were noted. Of the 716 bibliographic records, 64 (9%) were in AACR2R format. The remaining 652 bibliographic records (91%) were recataloged to conform to AACR2R. The Library of Congress name authority file in OCLC was searched to ensure that rules for choice among different names (e.g., pseudonyms, change of names) and forms of the same name (e.g., fullness, spelling) were not overlooked.

The steps in creating the cataloging record were retraced using AACR2R.¹¹ Each numbered rule used in this cataloging process was recorded; use of any instruction within a rule counted as a use of that rule. Multiple uses of rules for a single record were counted. For example, if a work required two added entries for collaborating persons, rule 21.30B1 was recorded as being used twice; if different parts of the rule were used, then more than one occurrence of the rule was recorded. For example, in a work of shared responsi-

bility with the principal responsibility indicated, rule 21.6B1 was recorded twice. The first part of the rule determines the main entry and the second part allows added entries. The following two rules were excluded in counting rule uses:

1. General rules for description in chapters 1 and 2 (i.e., 1.0-1.0H and 2.0-2.0H, rules that prescribe sources of information, organization of description, punctuation, language and script of the description, inaccuracies, accents and other diacritical marks, etc.); and
2. Introductory and general rules in chapter 21 (i.e., 21.0A-21.1B1, rules that prescribe the source for determining access points, definitions of works of personal authorship and corporate body).

In addition, rule 21.1B2 (entry under corporate body), which provides for entering "a work emanating from one or more corporate bodies under the heading for the appropriate corporate body" was applied judiciously.¹² The LC rule interpretation was followed. It states that a work emanates from a corporate body if that body

has issued (published) the work. Normally this means that the name of the corporate body appears in a position indicative of

publication (e.g., for books, the imprint position) in the chief source of information or appears elsewhere as a formal publication statement.¹³

This condition occurs in most, if not all, books. For this reason rule 21.1B2 was excluded from the count of rule uses except in cases where the corporate body functions other than as a commercial publisher.

FINDINGS

All books were categorized by patterns of authorship as follows:

Single authorship (personal name or corporate name) (57.8%) 414

Mixed responsibility or modifications (15.5%) 111

Shared responsibility (13.2%) 94

Editorial direction (11.3%) 81

Complex authorship (2.2%) 16

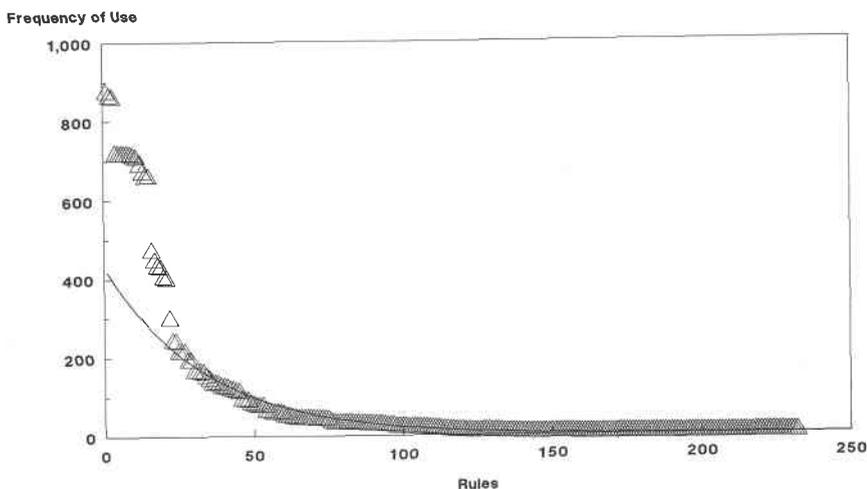
Total 716

Definitions for the first four categories were accepted as stated in AACR2R's glossary. The term *complex authorship* was coined to describe a condition of authorship where responsibility for the creation of a work resulted from a combination of two or more patterns of authorship (e.g., shared and mixed responsibilities or editorial direction and mixed responsibility).

More than half (57.8%) of the monographs in the practice cataloging collection have no more than one author, and the remaining 42.2% are quite evenly distributed among the various forms that have two or more authors. This corresponds with Svenonius, Baughman, and Molto's finding that "the profile of the typical monograph in the population of English language monographs currently received at a large research library may be characterized as having no more than one author/writer (61.25%)"¹⁴ The number of uses of all rules is given in appendixes to this report. Appendix A lists the rules in numerical order, with frequency of use by pattern of authorship. Appendix B lists rules by frequency of use in the entire collection.

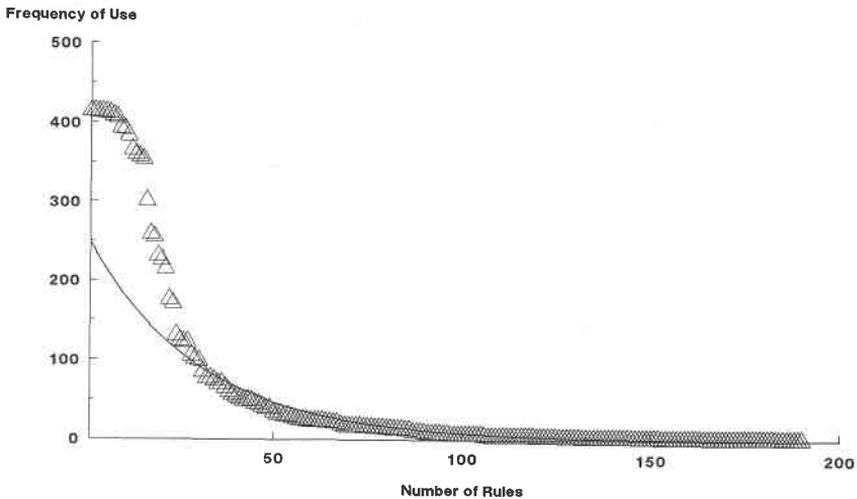
The frequency of use found in appendix B was plotted against the number of rules being used X times. The resulting chart (figure 3) shows that, for the collection as a whole, relatively few rules account for most of the uses. The curve of best fit is an exponential curve ($y=ae^{bx}$). When only works by a single person or corporate author are considered, the exponential curve again best describes the frequency of rule use (figure 4).

Seven chapters in AACR2R were used



Exponential curve
R-squared=.9744

Figure 3. All Rules Frequency of Use (n=716).



Exponential curve
R-squared=.9609

Figure 4. Single Person or Corporate Body (n=414).

in preparing bibliographic records for the 716 monographic publications studied, namely chapters 1, 2, 21, 22, 23, 24, and 25. Table 1 shows that only 232 (28.4%) of the 818 rules listed in these chapters were used. These in turn required a total of 20,247 rule uses. However, one must take into account that "AACR2 is based on the premise of a fully integrated library collection, with all types of materials being cataloged under the same rules and principles regardless of physical format."¹⁵ Thus, the code includes 188 rules (23%) for special types of materials and languages. Eliminating these reduces the number of potentially applicable rules to 615, and use increases to 38.2% of potentially applicable rules.

Meador and Wittig found use of "only 12 out of 143 rules listed in chapter 21 (approximately 8%) . . . for books in the economics sample" and "in the chemistry sample only 22 rules (approximately 15%)."¹⁶ In our study we found that 45 (31.5%) of the 143 rules in chapter 21 were used in determining access points. The differing results can be partially explained by the dissimilarity in the nature and scope of the collections examined. The Meador and Wittig samples were limited to economics and chemistry, while our proj-

ect was not limited to any particular discipline. In our study 30 of the 106 rules listed in chapter 22 (28.3%) were used, 28 (31.8%) of 88 rules included in chapter 24 were consulted, and 30 (22.6%) out of 133 rules in chapter 25 were used.

From chapters 21, 22, 24, and 25, 148 specialized rules can be removed from consideration because they deal with specifics for certain languages, courts, armed forces, embassies, intergovernmental and religious bodies, liturgical works, musical works, and sacred scriptures (except the Bible). Table 1 shows that even with these removals, 27% to 55% of the rules are used.

Table 1 shows a small change in the number of potentially applicable rules in chapters 1 and 2 when specialized rules are removed. Chapter 1 deals with general principles of bibliographic description applicable to all types of materials in all languages, while chapter 2 prescribes rules for bibliographic description of monographic materials, the major component of this collection. In addition, what would seem to be a cut-and-dried procedure in deciding rule use became considerably more complex because of overlapping rule usage. Within each numbered rule is an embedded subset of rules for recording

TABLE 1
NUMBER AND PERCENT OF AACR2R RULES USED BY CHAPTER

Chapter	No. of Rules	No. Used	%	
1. General Rules for Description	200	(187)	52	26.0 (27.8)
2. Books, Pamphlets, and Printed Sheets	133	(106)	44	33.1 (41.5)
21. Choice of Access Points	143	(125)	45	31.5 (36.0)
22. Headings for Persons	106	(73)	30	28.3 (41.1)
23. Headings for Geographic Names	15		3	20.0
24. Headings for Corporate Bodies	88	(70)	28	31.8 (40.0)
25. Uniform Titles	133	(54)	30	22.6 (55.6)
Total	818	(615)	232	28.4 (37.7)

The numbers enclosed in parentheses represents the total number of AACR2R rules when specific rules for certain languages and special types of materials are removed from consideration. The numbers of rules removed are chapter 1 (13), chapter 2 (27), chapter 21 (18), chapter 22 (33), chapter 24 (18), and chapter 25 (79).

information based on the "if" condition of the data element transcribed. A case in point is rule 1.1B1, which sets out four conditions of a title proper. Thus, simply citing and counting 1.1B1 does not reveal which of these four is the most frequently occurring condition for the works cataloged.

Table 2 shows the core set of rules for all monographs ranked by frequency of use. The core set refers to rules where the cumulative sum when arranged in descending order equals 90%. The 25 rules in the core are what a cataloger would expect: the three highest-frequency rule uses are those for choice of name of persons, and the two rules (22.1A and 22.1B) that have the highest frequency of use are rules that are first consulted whenever a personal author, regardless of function performed (e.g., writer, editor, translator, illustrator) in the creation of a work, is deemed necessary as an access point. Furthermore, this ranking supports the commonly accepted notion that individuals perform the function of writing. The core also suggests the predominance of entry under surname (22.5A1) over elements other than surname. The next 12 rules deal with information describing the item being cataloged. These are mostly data elements found on the title page. Be-

cause the cataloging of monographs is also governed by rules in chapter 1, there exists a one-to-one correspondence in rule citation between chapters 1 and 2 for identical data elements. For example, in rules 2.1B1 (title proper), 2.1F1 (statements of responsibility), 2.4C1 (place of publication), 2.4D1 (publisher), and 2.4F1 (date of publication), the cataloger is referred to another chapter for detailed instructions in transcribing the data elements. There is an abundance of these "swing to and fro" rules in AACR2R.

Rule 22.17A (addition of dates to distinguish identical names) was used frequently. This is because most bibliographic records in the study have forms of heading for personal names established when addition of dates (birth and/or death) was the rule rather than the exception (i.e., the practice of "no-conflict" cataloging had not been fully implemented by the Library of Congress).

The core set of rules (table 3) is from four chapters of the seven used in preparing bibliographic records for the 716 monographs. There are eight rules (4% of the rules in the chapter) from chapter 1 and 11 rules (8.3%) from chapter 2. Only two rules in chapter 21 appear in the core set, title added entries (21.30J1) and works of single personal authorship (21.4A1). The

TABLE 2
AACR2R CORE SET OF RULES RANKED BY FREQUENCY OF USE
(n = 25)

Rule No.	Rule Title	Frequency
	Headings for Persons	
22.1A	Choice of name—General rule	872
22.1B	Choice of name—General rule	861
22.5A1	Entry under surname—General rule	858
2.1B1	Title proper	716
1.1B1	Title proper	716
2.1F1	Statements of responsibility	716
2.4C1	Place of publication, distribution, etc.	716
2.4D1	Name of publisher, distributor, etc.	716
2.4F1	Date of publication, distribution, etc.	716
2.5D1	Dimensions	716
1.4D1	Name of publisher, distributor, etc.	710
1.4D2	Name of publisher, distributor, etc.	706
1.4C1	Place of publication, distribution, etc.	688
1.1F1	Date of publication, distribution, etc.	688
2.5B2	Number of volumes and/or pagination (Single volumes)	657
21.30J1	Added entries—Titles	656
22.17A	Additions to distinguish identical names—Dates	466
1.4F1	Date of publication, distribution, etc.	441
2.5C1	Illustrative matter	426
2.7B18	Notes (Contents: bibliographies, index, etc.)	423
2.5B1	Number of volumes and/or pagination (Single volumes)	398
21.4A1	Works of single personal authorship	395
2.1E1	Other title information	294
1.1E1	Other title information	294
1.4F6	Date of publication, distribution, etc.	238

presence of rule 21.4A1 in the core supports Meador and Wittig's observation of the "predominance of the rule for single authorship over all other rules."¹⁷

Table 4 shows the number of rules in the core set by pattern of authorship, which ranged from 14 to 19. There is no discernable difference in the number of rules used in the creation of bibliographic records by pattern of authorship. Almost identical rule numbers are cited in each category. The absence of rule 21.30J1 (title added

entries) for works prepared under editorial direction is conspicuous. This happens because in most cases main entry for works under editorial direction is under title.

The number of rules used in cataloging each book ranged from 15 to 54 (table 5), with the average for all books at 28.6 rules. The high numbers in the complex authorship category can be explained by the fact that in this category one is dealing with a number of possible combinations (e.g., shared and mixed; or editorial, shared, and

TABLE 3
AACR2R CORE SET OF RULES IN NUMERICAL ORDER FOR ALL BOOKS
(n = 25)

Rule No.	Rule Title	Frequency
1.1B1	Title proper	716
1.1E1	Other title information	294
1.1F1	Statements of responsibility	668
1.4C1	Place of publication, distribution, etc.	668
1.4D1	Name of publisher, distributor, etc.	710
1.4D2	Name of publisher, distributor, etc.	706
1.4F1	Date of publication, distribution, etc.	441
1.4F6	Date of publication, distribution, etc.	238
2.1B1	Title proper	716
2.1E1	Other title information	294
2.1F1	Statements of responsibility	716
2.4C1	Place of publication, distribution, etc.	716
2.4D1	Name of publisher, distributor, etc.	715
2.4F1	Date of publication, distribution, etc.	715
2.5B1	Number of volumes and/or pagination (Single volumes)	398
2.5B2	Number of volumes and/or pagination (Single volumes)	657
2.5C1	Illustrative matter	426
2.5D1	Dimensions	716
2.7B18	Notes (Contents: bibliographies, index, etc.)	423
21.4A1	Works of single personal authorship	395
21.30J1	Added entries—Titles	656
	Headings for Persons	
22.1A	Choice of name—General rule	872
22.1B	Choice of name—General rule	861
22.5A1	Entry under surname—General rule	858
22.17A	Additions to distinguish identical names—Dates	466

mixed) entailing use of more rules to describe the various combinations.

CONCLUSION

Fairthorne has observed that bibliometric distributions "arise from the way people choose to observe, arrange, and talk about things, not from the nature of things them-

selves."¹⁸ However, catalogers who create bibliographic records do not set out to use the cataloging rules in ways that would result in exponential or power curves. Rather, cataloging is similar to other complex human behaviors that, when observed, generate intriguing patterns or follow "bibliometric laws."

This study confirms earlier findings,

TABLE 4

AACR2R CORE SET OF RULES IN NUMERICAL ORDER BY PATTERN OF AUTHORSHIP

Single Authorship (n = 19)	Shared Authorship (n = 14)	Editorial Direction (n = 16)	Mixed Responsibility (n = 14)	Complex Authorship (n = 15)
1.1B1	1.1B1	1.1B1	1.1B1	1.1B1
1.1F1	1.1F1		1.1F1	1.1F1
1.4C1		1.4C1		
1.4D1	1.4D1	1.4D1	1.4D1	1.4D1
1.4D2		1.4D2	1.4D2	1.4D2
1.4F1				
2.1B1	2.1B1	2.1B1	2.1B1	2.1B1
2.1F1	2.1F1	2.1F1	2.1F1	2.1F1
2.4C1	2.4C1	2.4C1	2.4C1	2.4C1
2.4D1	2.4D1	2.4D1	2.4D1	2.4D1
2.4F1	2.4F1	2.4F1	2.4F1	2.4F1
2.5B2				
				2.5C1
2.5D1	2.5D1	2.5D1	2.5D1	2.5D1
2.7B18				
21.4A1				
	21.6A1			21.6A1
	21.6C1			
		21.7A1		
		21.7B1		
		21.8A1		21.8A1
		21.30D1		
21.30J1	21.30J1	21.30J1	21.30J1	21.30J1
22.1A	22.1A	22.1A	22.1A	22.1A
22.1B	22.1B	22.1B	22.1B	22.1B
22.5A1	22.5A1	22.5A1	22.5A1	22.5A1

TABLE 5

NUMBER OF AACR2R RULES USED PER BOOK

Pattern of Authorship	Avg. (Mean)	Minimum	Maximum
Single Authorship (n = 414)	25.8	15	54
Mixed Responsibility (n = 111)	30.2	22	48
Shared Authorship (n = 94)	33.5	23	54
Editorial Direction (n = 81)	33.1	21	52
Complex Authorship (n = 16)	36.9	30	51
All (n = 716)	28.6	15	54

namely the existence of a skewed distribution (few rules cover almost all conditions in the preparation of a bibliographic record) and the existence of a core set of rules sufficient in most cases to describe any given item. Several observations, including this study, have found that frequency of cataloging rule use is best described by an exponential curve; the power curve that fits Fox's analysis of the ALA and AACR rules for main entry is an extreme form of the exponential curve. The *Anglo-American Cataloguing Rules* in their various editions produce remarkably similar distributions of use, whether one studies the entire descriptive cataloging process or focuses on a component such as choice of access points. It is also clear that most users need to know a relatively small core set of rules to be able to understand the nature of a bibliographic record. Concentrating on this core could be part of what Richmond envisioned as a "simple step [to] mollify many of those who criticize the catalog as being too hard to use."¹⁹

The implications for instruction of new catalogers and development of expert systems are clear. It is reassuring that empirical analysis confirms the cataloger's common sense: rules for name, title, statement of responsibility, place of publication, name of publisher, and date of publication are common, essential elements of the bibliographic record. A relatively small and stable core of cataloging rules is consistently used, and education should begin with and focus on this core. The rapid decrease in number of uses for many of the rules suggests that most catalog users and expert systems should focus on the core set of rules, safely ignoring those less frequently used. In a way, this empirical study upholds Osborn's call in the early 1940s for a practical rather than legalistic approach to the "crisis in cataloging." Faced with overwhelming complexity in rules and interpretations, Osborn suggested that "rules for cataloging would be relatively few and simple."²⁰ We now have evidence that most books indeed can be cataloged with a set of pragmatically derived core rules.

REFERENCES

1. Dorothy Gregor and Carol Mandel, "Cataloging Must Change!" *Library Journal* 116, no.6:42-47 (April 1, 1991).
2. Roy Meador III and Glenn R. Wittig, "AACR2 Rules Used in Assigning Access Points for Books in Two Subjects: Implications for Automatic Cataloging Expert Systems," *Library Resources & Technical Services* 35:135-40 (1991).
3. Roland Hjerpe and Birgitta Olander, "Cataloging and Expert Systems: AACR2 as a Knowledge Base." *Journal of the American Society for Information Science* 40:27-44 (1989).
4. Roy Meador III and Glenn R. Wittig, "Expert Systems for Automatic Cataloging Based on AACR2: A Survey of Research." *Information Technology and Libraries* 7:168 (1988).
5. Ling-Hwey Jeng, "An Expert System for Determining Title Proper in Descriptive Cataloging: A Conceptual Model," *Cataloging & Classification Quarterly* 7, no.2:55-70 (Winter 1986).
6. Roland Hjerpe and Birgitta Olander, *Artificial Intelligence and Cataloging: Building Expert Systems for Simple Choice of Access Points; Results and Revelations* (Linköping, Sweden: LIBLAB, Department of Computer and Information Science, Linköping University, 1985).
7. Meador and Wittig, "AACR2 Rules."
8. Roy Davies, "Cataloging as a Domain for an Expert System," in *Intelligent Information Systems: Progress and Prospects*, ed. R. Davies, Ellis Horwood Series in Artificial Intelligence (New York: Halstead, 1986), p.69.
9. Ann Sandberg Fox, *The Amenability of a Cataloging Process to Simulation by Automatic Techniques* Ph.D. diss., Univ. of Illinois at Urbana-Champaign, 1972.
10. Phyllis A. Richmond, "Classification from PRECIS: Some Possibilities," *Journal of the American Society for Information Science* 27:240-47 (1976).
11. Statistical analysis including frequency counts and curve fitting were accomplished using the *CSS Statistica* and *Lotus Graphwriter II* software packages.
12. *Anglo-American Cataloguing Rules, Second Edition, 1988 Revision*, prepared under the direction of the Joint Steering Committee for Revision of AACR, ed. Michael Gorman and Paul W. Winkler (Chicago: American Library Assn., 1988), p.313.
13. *Library of Congress Rule Interpretations*,

- 2d ed. (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1989), p. 21.1B2, page 1.
14. Elaine Svenonius, Betty Baughman, and Mavis Molto, "Title Page Sanctity? The Distribution of Access Points in a Sample of English Language Monographs," *Cataloging & Classification Quarterly* 6, no. 3:15 (Spring 1986).
 15. Margaret F. Maxwell, *Handbook for AACR2 1988 Revision: Explaining and Illustrating the Anglo-American Cataloguing Rules* (Chicago: American Library Assn., 1989), p.20.
 16. Meador and Wittig, "AACR2 Rules," p.139.
 17. Meador and Wittig, "AACR2 Rules," p.139.
 18. Robert A. Fairthorne, "Empirical Hyperbolic Distributions (Bradford-Zipf-Mandelbrot) for Bibliometric Description and Prediction," *Journal of Documentation* 25:340 (1969).
 19. Phyllis A. Richmond, "Some Theoretical Considerations," *American Libraries* 6:606 (1975).
 20. Andrew D. Osborn, "The Crisis in Cataloging," *Library Quarterly* 11:404 (1941).

APPENDIX A
RULES IN NUMERICAL ORDER

Rule Number	Uses	Rule Number	Uses	Rule Number	Uses
1.1B1	48	1.4F8	15	2.7B5	6
1.1B2	9	1.5E1	1	2.7B6	41
1.1D1	1	1.6B1	211	2.7B7	33
1.1D2	1	1.6D1	1	2.7B9	45
1.1E1	15	1.6E1	2	2.7B10	81
1.1E3	6	1.6C	117	2.7B12	2
1.1E6	5	1.6H1	4	2.7B13	2
1.1F1	668	1.6H2	3	2.7B14	2
1.1F2	57	1.6H5	3	2.7B17	30
1.1F3	41	1.6J1	1	2.7B18	423
1.1F4	112	2.1B1	716	2.7B19	31
1.1F5	17	2.1D1	10	2.8C1	28
1.1F6	189	2.1E1	294	21.1B2	121
1.1F7	189	2.1F1	715	21.1B3	57
1.1F8	5	2.2B1	150	21.1C1	48
1.1F12	1	2.2C1	1	21.4A1	395
1.1F13	1	2.2E1	1	21.4B1	61
1.1F15	30	2.4C1	716	21.5A	4
1.1G1	1	2.4D1	715	21.5B	1
1.1G2	8	2.4F1	715	21.6A1	110
1.1G3	8	2.5B1	398	21.6B1	31
1.2B1	132	2.5B2	657	21.6B2	31
1.2B4	19	2.5B3	211	21.6C1	134
1.2C1	6	2.5B5	7	21.6C2	24
1.2E1	1	2.5B7	31	21.7A1	78
1.4C1	688	2.5B8	3	21.7B1	142
1.4C3	122	2.5B10	40	21.7C1	17
1.4C5	159	2.5B13	1	21.8A1	130
1.4C6	29	2.5B17	44	21.9A	42
1.4D1	710	2.5C1	426	21.10A	10
1.4D2	706	2.5C2	115	21.11A1	88
1.4D3	41	2.5C3	58	21.12A1	20
1.4D4	25	2.5C4	5	21.13A1	1
1.4D5	18	2.5C5	39	21.13C1	2
1.4D6	2	2.5D1	713	21.14A	28
1.4D7	4	2.5D2	4	21.17B1	2
1.4E1	4	2.5E1	1	21.19A1	2
1.4F1	441	2.6B1	211	21.20A1	1
1.4F2	4	2.7B1	42	21.24A	2
1.4F5	74	2.7B2	20	21.29C	61
1.4F6	238	2.7B3	10	21.30A1	75
1.4F7	24	2.7B4	26	21.30B1	6

APPENDIX A CONTINUED
RULES IN NUMERICAL ORDER

Rule Number	Uses	Rule Number	Uses	Rule Number	Uses
21.30C1	125	22.8B1	1	24.21A	4
21.30D1	163	22.12B1	6	24.21B	3
21.30E1	163	22.12B1	1	24.21C	1
21.30F1	7	22.15B1	2	24.26A	1
21.30G1	20	22.15C	56	25.1A	42
21.30H1	2	22.16D1	1	25.2A	42
21.30J1	656	22.17A	466	25.2C1	4
21.30K1	20	22.18A	89	25.2E1	19
21.30K2	17	22.19B1	28	25.2E2	13
21.30L1	91	23.2A1	2	25.3A	28
21.31A1	3	23.4B1	1	25.4A1	5
21.31B1	5	23.4E1	1	25.4C1	3
21.32A1	1	24.1A	237	25.5B1	1
21.33A	2	24.1C1	3	25.5C1	13
21.37A	9	24.2B	5	25.6B3	4
22.1A	872	24.2C	4	25.8A	2
22.1B	861	24.2D	7	25.9A	2
22.1C	3	24.3A1	2	25.15A1	3
22.1D1	11	24.3B1	3	25.15A2	3
22.1D2	3	24.4C2	19	25.17A	9
22.2A1	16	24.4C3	5	25.18A1	3
22.2B1	8	24.4C4	3	25.18A2	3
22.2B2	2	24.5A1	42	25.18A3	2
22.2B3	1	24.5C1	15	25.18A8	2
22.2C1	5	24.7A1	8	25.18A9	1
22.3A1	161	24.7B1	26	25.18A10	8
22.3B2	1	24.7B2	8	25.18A11	6
22.3C1	1	24.7B3	27	25.18A13	6
22.3D1	1	24.7B4	26	25.18M1	1
22.5A1	858	24.12A	27	25.25A	1
22.5C1	1	24.13A	26	25.26A	1
22.5C2	1	24.14A	23	25.27A1	1
22.5C3	11	24.17A	74	25.35E1	1
22.5D1	13	24.18A	53	25.35F1	1
22.6A1	2	24.19A	42		
22.8A1	7	24.20B1	1		

APPENDIX B
RULES IN ORDER OF FREQUENCY OF USE

Uses	Rule Number	Uses	Rule Number	Uses	Rule Number
872	22.1A	426	2.5C1	142	21.7B1
861	22.1B	423	2.7B18	134	21.6C1
858	22.5A1	398	2.5B1	132	1.2B1
716	2.1B1	395	21.4A1	130	21.8A1
716	2.4C1	294	2.1E1	125	21.30C1
715	2.1F1	238	1.4F6	122	1.4C3
715	2.4D1	237	24.1A	121	21.1B2
715	2.4F1	211	1.6B1	117	1.6G
713	2.5D1	211	2.5B3	115	2.5C2
710	1.4D1	211	2.6B1	112	1.1F4
706	1.4D2	189	1.1F6	110	21.6A1
688	1.4C1	189	1.1F7	91	21.30L1
668	1.1F1	163	21.30D1	89	22.18A
657	2.5B2	163	21.30E1	88	21.11A1
656	21.30L1	161	22.3A1	81	2.7B10
466	22.17A	159	1.4C5	78	21.7A1
441	1.4F1	150	2.2B1	75	21.30A1

Uses	Rule Number	Uses	Rule Number	Uses	Rule Number
74	1.4F5	15	24.5C1	3	25.18A1
74	24.17A	13	22.5D1	3	25.18A2
61	21.4B1	13	25.2E2	2	1.4D6
61	21.29C	13	25.5C1	2	1.6E1
58	2.5C3	11	22.1D1	2	2.7B12
57	1.1F2	11	22.5C3	2	2.7B13
57	21.1B3	10	21.10A	2	2.7B14
56	22.15C	10	2.1D1	2	21.13C1
53	24.18A	10	2.7B3	2	21.17B1
48	1.1B1	9	1.1B2	2	21.19A1
48	21.1C1	9	21.37A	2	21.24A
45	2.7B9	9	25.17A	2	21.30H1
44	2.5B17	8	1.1C2	2	21.33A
42	2.7B1	8	1.1G3	2	22.2B2
42	21.9A	8	22.2B1	2	22.6A1
42	24.5A1	8	24.7A1	2	22.15B1
42	24.19A	8	24.7B2	2	23.2A1
42	25.1A	8	25.18A10	2	24.3A1
42	25.2A	7	2.5B5	2	25.8A
41	2.7B6	7	21.30F1	2	25.9A
41	1.1F3	7	22.8A1	2	25.18A3
41	1.4D3	7	24.2D	2	25.18A8
40	2.5B10	6	1.1E3	1	1.1D1
39	2.5C5	6	1.2C1	1	1.1D2
33	2.7B7	6	2.7B5	1	1.1F12
31	2.5B7	6	21.30B1	1	1.1F13
31	2.7B19	6	22.12B1	1	1.1G1
31	21.6B1	6	25.18A11	1	1.2E1
31	21.6B2	6	25.18A13	1	1.5E1
30	1.1F15	5	1.1E6	1	1.6D1
30	2.7B17	5	1.1F8	1	1.6J1
29	1.4C6	5	2.5C4	1	2.2C1
28	2.8C1	5	21.31B1	1	2.2E1
28	21.14A	5	22.2C1	1	2.5B13
28	22.19B1	5	24.2B	1	2.5E1
28	25.3A	5	24.4C3	1	21.5B
27	24.7B3	5	25.4A1	1	21.13A1
27	24.12A	4	1.4D7	1	21.20A1
26	2.7B4	4	1.4E1	1	21.32A1
26	24.7B1	4	1.4F2	1	22.2B3
26	24.7B4	4	1.6H1	1	22.3B2
26	24.13A	4	2.5D2	1	22.3C1
25	1.4D4	4	21.5A	1	22.3D1
24	1.4F7	4	24.2C	1	22.5C1
24	21.6C2	4	24.21A	1	22.5C2
23	24.14A	4	25.2C1	1	22.8B1
20	2.7B2	4	25.6B3	1	22.16D1
20	21.12A1	3	1.6H2	1	23.4B1
20	21.30G1	3	1.6H5	1	23.4E1
20	21.30K1	3	2.5B8	1	24.20B1
19	1.2B4	3	21.31A1	1	24.21C
19	24.4C2	3	22.1C	1	24.26A
19	25.2E1	3	22.1D2	1	25.5B1
18	1.4D5	3	24.1C1	1	25.18A9
17	1.1F5	3	24.3B1	1	25.18M1
17	21.7C1	3	24.4C4	1	25.25A
17	21.30K2	3	24.21B	1	25.26A
16	22.2A1	3	25.4C1	1	25.27A1
15	1.1E1	3	25.15A1	1	25.35E1
15	1.4F8	3	25.15A2	1	25.35F1

Bibliographic Relationships: An Empirical Study of the LC Machine-Readable Records

Barbara B. Tillett

In 1986 an empirical study was conducted to examine the extent of bibliographic relationships as reflected in their frequency of occurrence within the 1968–July 1986 machine-readable database of the Library of Congress. Frequency of occurrence was determined by counting the incidences of specific codes associated with each relationship type within the machine-readable records. Also examined were characteristics of bibliographic items exhibiting particular relationships. Characteristics of interest were language, place of publication, publication date, subject, and bibliographic format, as it was thought such factors might prove useful in predicting particular types of relationships for future cataloging systems. Such information can be of potential use to decision makers and system designers in assessment of appropriate methods for designating specific relationships in both future catalogs and future cataloging rules.

Library catalogs traditionally have identified relationships among bibliographic items represented. These relationships can be categorized into the following:

1. equivalence relationships, which hold between exact copies of the same manifestation of a work or between an original item and reproductions of it, as long as intellectual content and authorship are preserved;
2. derivative relationships, which hold between a bibliographic item and a modification based on that item;
3. descriptive relationships, which hold

- between a bibliographic item or work and a description, criticism, evaluation, or review of that item or work;
4. whole-part relationships, which hold between a component part of a bibliographic item or work and its whole;
5. accompanying relationships, which hold between a bibliographic item and the bibliographic item it accompanies, such that the two items augment each other equally or one item augments the other principal or predominant item;
6. sequential relationships, which hold between bibliographic items that con-

BARBARA B. TILLET is Head of the Catalog Department, University of California, San Diego. Invited paper received and accepted for publication December 16, 1991. (Note: Figures 3–9 and portions of figure 1 reproduced with permission of author and IFLA.) *Editor's Note:* Library Resources & Technical Services presents the fourth and final research report on the topic of bibliographic relationships. These reports are derived from the author's 1987 Ph.D. dissertation, "Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Relationships Used in Cataloging."

tinue or precede one another, but are not considered derivative; and

7. shared characteristic relationships, which hold between a bibliographic item and another bibliographic item that is not otherwise related but coincidentally has a common author, title, subject, or other characteristic used as an access point in a catalog.

The devices used to link records have varied over the years and have been greatly influenced by the type of catalog available. To take one example, card catalogs facilitated the easy creation of "added entries," which allowed the display of full bibliographic information under each heading in the catalog and increased collocation of various editions of works, thereby linking records for bibliographic items with derivative relationships. Online catalogs and future computerized catalogs may offer us opportunities to develop even more powerful and useful linking devices to help the catalog user navigate.

Assuming the usefulness of linking related bibliographic records in the library catalog, what portion of the records are we talking about? For each type of relationship, how many records can we expect will need to be linked and how extensive will the links be?

To answer those questions partially, an empirical study was conducted to examine the extent of bibliographic relationships as reflected in their frequency of occurrence within the 1968–July 1986 machine-readable database of the Library of Congress. Frequency of occurrence was determined by counting the incidences of specific codes associated with each relationship type within the machine-readable records. Also examined were characteristics of bibliographic items exhibiting particular relationships. Characteristics of interest were language, place of publication, publication date, subject, and bibliographic format, as it was thought such factors might prove useful in predicting particular types of relationships for future cataloging systems. The general count of occurrence of *relationships* and the count of occurrence of *characteristics* tells us approximately how many and what kind of bibliographic records exhibit a particular

type of relationship. Such information can be of potential use to decision makers and system designers in assessment of appropriate methods for designating specific relationships in both future catalogs and future cataloging rules.

POPULATION CHARACTERISTICS

The population for the empirical study was the set of bibliographic records in the 1968–July 1986 machine-readable database of the Library of Congress. The LC database included library materials cataloged by the Library of Congress since 1968, when the library's machine-readable bibliographic records were first created. In the early years the database included only English-language monographs, but since 1976 other roman-alphabet languages have been included; transliterated nonroman languages were added later. The database as of July 1986 included nearly all languages and nearly all types of bibliographic items: books, serials,¹ audiovisual materials, music, and maps. As such, it included all of the products of general publishing.² The only types of material excluded were manuscripts and "machine-readable data files" (or computer files); however, counts from the OCLC bibliographic database, which includes manuscripts and computer files, indicated that as of 1986, manuscripts accounted for less than 0.4% and computer files for 0.08% of bibliographic records in machine-readable form.³ In other words, manuscripts or computer files were the least prevalent of all types of materials; thus the LC database, though it excluded bibliographic records for them, could still be regarded as comprehensive and essentially representative of bibliographic databases in major research libraries.

The population size, i.e., the number of records in the 1968–July 1986 LC machine-readable database, was 2,854,252. The number of machine-readable records for each bibliographic format of material varied widely. For example, the number of records in the database ranged from approximately 21,000 records for music to more than 2 million records for books. Figure 1 and table 1 reflect the database

size during this study. The dates after each material type indicate when material was cataloged and added to the database.

COUNTING THE FREQUENCY OF OCCURENCE OF RELATIONSHIPS

Frequency of occurrence was determined by counting the incidences of specific codes or text strings associated with each relationship type. Such a count tells us the frequency of occurrence of the specified code representing a type of relationship in the entire database. The codes used were the machine-readable codes in the MARC formats used for bibliographic description. These machine-readable codes include *tags*, *indicators*, *subfield codes*, and for some fields, coded *values*. In the MARC format, *tags* are three-digit numeric codes, while indicators and subfield codes are single-digit alphanumeric codes. Coded values are typically of one to three digits. Take the example of the MARC code

041 1 #a eng #h ger
Here

- the "041" tag identifies a language

- field in a bibliographic record;
- the indicator "1" means the described bibliographic item is or includes a translation;
- the subfield code "#a" identifies the language of the text, and the subfield code "#h" identifies the language of the original from which the translation was made; and
- the values "eng" and "ger" specify the

TABLE 1

NUMBER OF RECORDS IN EACH MARC FILE AT THE LIBRARY OF CONGRESS

MARC File	No. of Records	% of Total
Books (1968-7/31/86)	2,330,074	81.6
Serials (1973-7/31/86)	321,646	11.3
Maps (1973-6/30/86)	101,408	3.5
Visuals (1972-6/30/86)	79,275	2.8
Music (4/84-6/30/86)	21,849	0.8
Total	2,854,252	100.0

Library of Congress Database, 1968-July 1986.

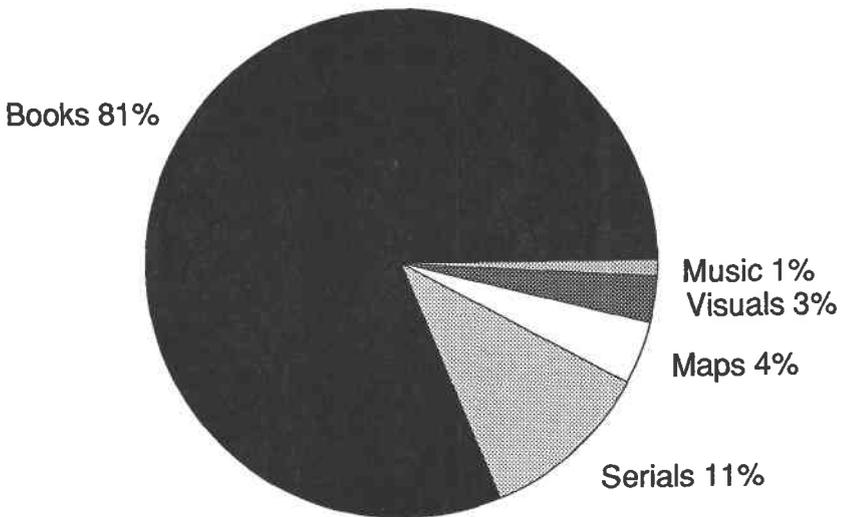


Figure 1. Library of Congress MARC Database.

particular languages as English and German.

Some relationships can be explicitly identified by a unique machine-readable code. In those cases it is not difficult to count the relationships in the entire population, because there is a one-to-one correspondence between the unique code and the relationship.⁴ However, some relationships are expressed in a field identified by a general code that encompasses more than one type of relationship: that is, the code has a one-to-many correspondence to relationships. Still other relationships have no associated code.

The general MARC codes, in addition to encompassing more than one type of relationship, may be used for more than relationship information, such as information about a corporate body identified on the title page above the title. Computer counting, which is easily accomplished for specific unique codes, would require a complex procedure to capture relationship information within fields with ambiguous general codes. The only general MARC code of relevance to this study is the "500" code, which tags a general note. About half of the LC database, or 1,380,000 records, have "500" general notes.⁵ Not only do the "500" general notes include both relationship and nonrelationship information, but there may also be several such notes within a given bibliographic record. As a result, determining the presence of information regarding any given relationship using the "500" tag requires scanning each record to find words or phrases associated with each bibliographic relationship, and then finally counting the occurrence of fields containing the identifying word or phrase. The cost of conducting a scan and count of "500" fields for the entire database was prohibitive. Consequently, a separate study on a sample of records with "500" general notes was conducted both to determine the types of bibliographic relationships included and to observe any patterns in the words that would uniquely identify a particular type of bibliographic relationship. The methodology and findings from the study of "500" general notes are discussed below under "STUDY OF '500,' GENERAL NOTES."

For the situation where no code is available to identify a given relationship, the relationship can be assumed to be reciprocal to a coded relationship.⁶ Examples of reciprocal relationships are those of a book to its review or an edition to its succeeding edition. These reciprocal relationships currently *are not* noted on bibliographic records, whereas relationships going in the reverse direction *are* noted through a citation to the earlier bibliographic item, for both reviews and editions. Often reciprocal relationships are accommodated in the catalog through collocating main entries and added entries. For example, for a later edition an added entry is made for the earlier edition which files next to the main entry for that earlier item, thereby collocating the records under the heading for the earlier record. The existence of codeless relationships of this reciprocal type in the population can be inferred from the presence of the coded reciprocal. Codeless relationships were included in the taxonomy of bibliographic relationships but were not included in this empirical study.⁷

Briefly, the empirical study was conducted in two parts: a study of those fields in which relationships are explicitly coded and a study of the "500" general notes fields in which both relationship and nonrelationship information is embedded. We will now turn our attention to the first of these two parts.

STUDY OF EXPLICITLY CODED RELATIONSHIPS: METHODOLOGY

The MARC formats were reviewed to select fields, other than the "500" fields, containing explicit bibliographic relationship information. The result was a target list of tags, indicators, subfield codes, and values totaling 134 specifically coded bibliographic relationships.⁸ Each relationship was assigned to an appropriate taxonomy category: equivalence, derivative, descriptive, whole-part, accompanying, sequential, and shared characteristic. As it turned out, there were no specific MARC codes other than the "500" general notes for the category of descriptive relationships. The category of shared characteris-

tic relationships, on the other hand, could be indicated by nearly every MARC code and was dropped from the empirical study due to the complexity of counting them.⁹ There were no records for 14 of the 134 coded relationships.¹⁰ The remaining 120 relationship codes were translated into computer queries by Leo Settler of the Library of Congress. A computer query is a statement of parameters for retrieving database records. An example is the query for retrieving all records in the MARC book file that have a "300" field with subfield "e." The set of retrieved records for a query can be considered a file of relationship records, and we will call the total set of records retrieved for all the queries the relationship file. A query is run against a set of programs to determine the number of retrieved records containing specific values of the factors subject, publication date, language, and country of publication.¹¹ Additionally, the researcher obtained LC's internal statistics, called "Descriptive Tabulations," which include frequency of occurrence data for each MARC code within each MARC format.

Because the Library of Congress has the *JANUS* system of computer programs for its own routine database management, a system that counts the frequency of occurrence of specific fields and subfields for each of its MARC files, we were able to run the queries to determine the number of occurrences of each relationship for each of the specific factors under study.

FACTORS

Relationships were analyzed in terms of various factors that were expected to affect their frequency of occurrence, namely,

1. bibliographic format (i.e., each MARC format for books, serials, maps, visuals, and music)
2. subject, defined in terms of ranges of LC classification numbers,
3. publication date,
4. language, and
5. country of publication.¹²

The hypothesis was that these factors would reflect distinctive patterns for each type of bibliographic relationship, lending themselves to prediction of a given rela-

tionship. It was thought that for each type of bibliographic relationship patterns might exist that could be used in cataloging rules or catalog systems design. With regard to cataloging rules, for example, if one could show that there were distinctive variations in bibliographic relationships exhibited by bibliographic format, this would help make a case for preferring a code with specific rules for specific materials rather than an integrated cataloging code. Similarly, if there were variations by subject, a special library might wish to design its online catalog to favor the predominant relationships for its subject specialty. In fact all of these factors could be used when defining any individual library's collection and could prove useful in the design of a catalog for that collection. The results of this aspect of the study are examined under the discussion of findings and in the summary.

The methodology used to analyze these factors varied with the factor. Fixed-field codes were used to identify publication date, language, and country of publication. The occurrences of each publication date were counted, and then the counts were grouped by decade for the twentieth century and by century for the nineteenth and eighteenth centuries; all pre-1700 dates were grouped together. The occurrences of each language code were tabulated; counts were separately noted for English, French, German, Italian, Portuguese, Russian, and Spanish, with all other languages grouped together as one count. Similarly, the occurrences for each country of publication code were grouped into twenty categories (see figure 9). Subject categories were determined from the call number field (tag 050) and grouped before counting into four broad areas based on the Library of Congress classification as follows:

- Sciences = GC, Qs, Rs, Ss, Ts, Us, Vs
- Social sciences = BF, Gs except GC, Hs, Js, Ks, Ls
- Humanities = Bs except BF, Cs, Ds, Es, Fs, Ms, Ns, Ps
- General = As and Zs, which cover all disciplines

A fifth value of "unknown" was added to the subject variable to account for bib-

TABLE 2
SOURCES FOR FACTORS STUDIED

Factor	Source
1. MARC format	the separate LC MARC files
2. Subject (LC classification)	050 field
3. Publication date	Dates (fixed-field code)
4. Language	Language (fixed-field code)
5. Country of publication	Country (fixed-field code)

liographic records without call number fields or with call number values other than Library of Congress classifications. (This is described further under the discussion of findings.)

Table 2 identifies the MARC record sources of the five factors studied. The corresponding figure 2 is a sample MARC record with each of the sources circled.

DATA ANALYSIS

As noted, each bibliographic relationship was translated into a query. For the results of each query, a data-collection form was completed in which counts of frequencies of occurrence for each value of the five factors (as reported below in the discussion of findings) were compiled. The frequen-

cies of occurrence thus recorded were then keyed into *SuperCalc* files. In order to compare bibliographic records incorporating relationship information to those in the LC database as a whole, the number of occurrences of fields encoding bibliographic relationships was translated into the number of records incorporating relationship information, which we will call relationship records, using the ratios of occurrences per record provided in LC's "Descriptive Tabulations." For example, the ratio of occurrences to records for a main series relationship (MARC field 760) is 1.10; the number of occurrences retrieved, 14,780, divided by 1.10, equals 13,436 records.¹³ The resulting tables of distribution of types of relationships by each factor over relationship records were printed and the corresponding graphs produced with *SuperCalc* software.

STUDY OF EXPLICITLY CODED RELATIONSHIPS: DISCUSSION OF FINDINGS

EXAMINATION OF LC MARC RECORDS

Three conditions complicated the analysis of results: the counting method, the limitations of LC's computers, and the nature of bibliographic records.¹⁴ Regarding the nature of bibliographic records, many bibliographic relationships can occur within one bibliographic record. For example, an

```

-----
Type: a   Bib lvl: m   Rec stat: n   Entrd: 701028   Used: 841109
Repr: a   Enc lvl:   Govt pub:   (Lang: eng)   Source:   Illus: ac
Indx: 1   Mod rec:   Conf pub: 0   (Ctry: enk)   Dat tp: s   M/F/B: 10
Desc:    Int lvl:   Festschr: 0   Cont: b
          010     77-118840
          040     DLC +c DLC
          (050 0   QD321 +b .A78 1970)
          082     547/.782
          100 10   Aspinall, Gerald O.
          245 0   Polysaccharides, +c by Gerald O. Aspinall.
          250     [1st ed.]
          260 0   Oxford, +a New York, +b Pergamon Press +c [1970]
          300     xvi, 228 p. +b ill., port. +c 20 cm.
          490 0   The Commonwealth and international library. A Course
                   in organic chemistry
          504     Includes bibliographical references.
          650 0   Polysaccharides.
-----

```

Note: This record is presented in the OCLC (First System) display format.

Figure 2. Sample MARC Record from the Book File (Circled Fields Correspond to Table 1 Factors).

item can have a microform copy, a revised edition, and a supplement, and can be an anthology of short stories, all of which are reflected in the bibliographic record for the item. Additionally, any given bibliographic record can include more than one instance of a given type of relationship, such as a record for a bibliographic item with an earlier edition and a later edition, which thereby has two derivative relationships to different items. Therefore, the count of records associated with any relationship type may include a given bibliographic record more than once. This prohibits our making a one-to-one correspondence between occurrences of relationships and occurrences of bibliographic records exhibiting those relationships. However, we can determine the number of records having a given bibliographic relationship from the occurrences of the bibliographic relationship using the ratios of occurrences per record.

As for the counting method, the data were not collected in a manner that would allow us to correlate factors pertaining to a particular bibliographic record, but rather were collected for each factor separately for each group of bibliographic records with a given code. Therefore, we do not have a base for statistically correlating data across factors. We can compare the factors, but cannot conduct statistical correlations between them.

The third condition complicating the analysis of results was a limitation on the use of the LC computer. Basically, the Library of Congress computer was running the computer queries for this study while conducting its routine operations. Exclusive use of the machine was not possible and meant that one query that involved very large numbers of retrieved records (that for the "440" series field) could not be conducted.¹⁵

The following sections examine five of the taxonomic categories of bibliographic relationships—equivalence, derivative, whole-part, accompanying, and sequential—in terms of the five factors—bibliographic format, subject, date of publication, language, and country of publication. As a reminder to the reader, "descriptive relationships" are only found in the "500" general notes and are discussed in a later

portion of the study. Also, the category of shared characteristics relationships is excluded from the study altogether due to the complexity of examining such relationships.

EXAMINATION OF FACTORS

MARC FORMATS

The number of occurrences of each relationship type found in the LC database was counted for each MARC format category: books, serials, maps, visual materials, and music. Not all records in the LC database exhibit bibliographic relationships. The percentage of bibliographic records exhibiting relationships as viewed across each format was compared to the distribution for the LC database as a whole in the graphic presentation in figure 3 and the corresponding table 3.

As expected, the distribution of MARC formats in the relationship file basically follows the distribution of MARC formats in the LC database as a whole. In both the LC database and the relationship file, most records are for books, followed by records for serials, maps, visual materials, and music. However, when we compare the percentage of records in the LC database as a whole for various MARC formats with the percentage of records in the relationship file for the various MARC formats, we notice a striking difference, as shown in table 3.

As shown in figure 3 and table 3, for the total records exhibiting relationships there are over three times as many maps as would be expected given the frequency in the LC database as a whole; nearly twice as many serials, visual materials, and music; and only a little more than three-quarters as many books. When we examine the distributions within each of the relationship types, we find even greater differences. For example, in records exhibiting accompanying relationships, visual materials are represented nearly twenty-four times more than for the database as a whole. Sequential relationships were found only in serial records. From these figures we could predict that visual materials and music would likely be involved in accompanying relationships, serials would

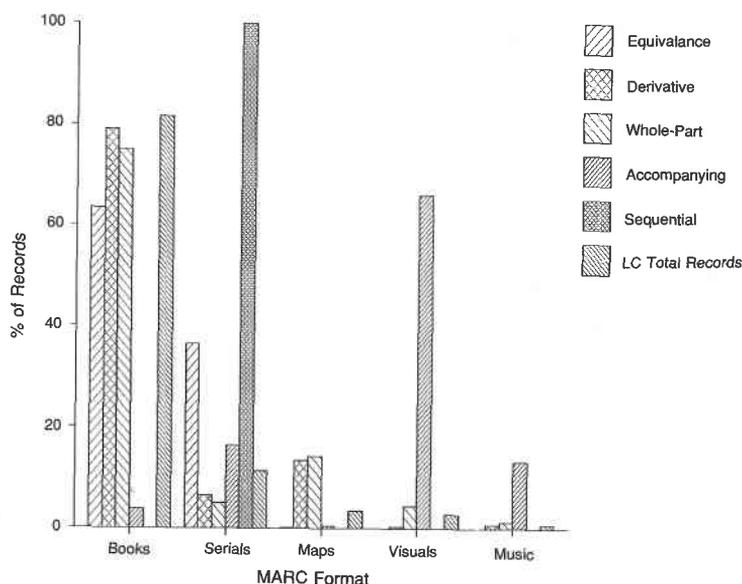


Figure 3. Distribution of MARC Formats over the Relationship Records (Percentage of Records).

TABLE 3
PERCENTAGE OF BIBLIOGRAPHIC RECORDS
DISPLAYING RELATIONSHIPS BY MARC FORMAT

Relationship Type	Books	Serials	Maps	Visuals	Music	Total
Equivalence	63.34	36.53	0.11	0.00	0.01	99.99
Derivative	79.03	6.46	13.41	0.33	0.77	100.00
Whole-Part	75.01	4.95	14.28	4.48	1.28	100.00
Accompanying	3.77	16.29	0.54	65.98	13.42	100.00
Sequential	0.00	100.00	0.00	0.00	0.00	100.00
Total	60.31	22.01	10.84	5.42	1.43	100.01
LC Database as a Whole	81.60	11.30	3.50	2.80	0.80	100.00

likely be involved in sequential relationships, while maps would likely be involved in derivative or whole-part relationships. Books would only rarely exhibit accompanying or sequential relationships.

In figure 4, we can clearly see the difference among the categories of MARC formats where we note that books are the most common format exhibiting equivalence, derivative, and whole-part relationships; visual materials predominate in

the accompanying category; and serials are the exclusive format in the sequential category.¹⁶ It must be remembered that these results do not include the relationships embedded in the general notes, which are examined in the separate study described below.

It is evident from figure 4 that the whole-part relationship category predominates for every type of material in every format except serials. For books alone the

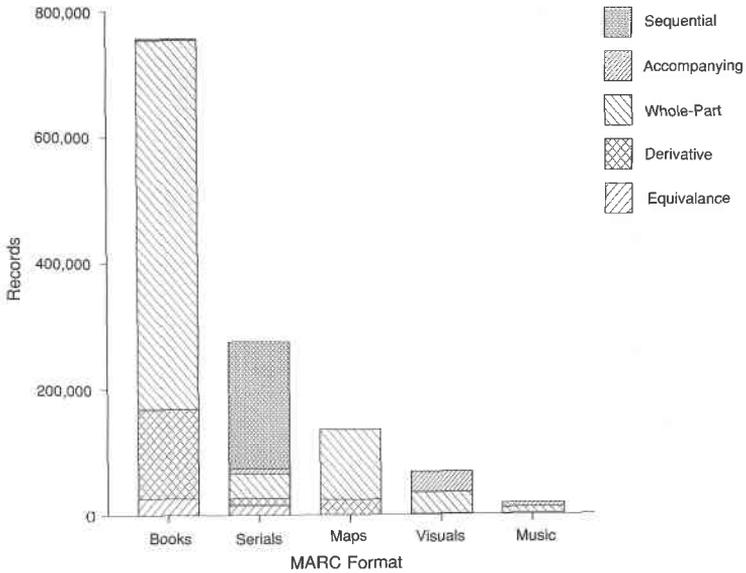


Figure 4. Distribution of Relationship Type by MARC Format.

whole-part relationship constitutes 75.01% of the records in the relationship file. The next most common bibliographic relationship category for books is the derivative relationship, such as that which occurs between editions or translations; these relationships appear in 18.72% of all records in the relationship file. For serials, sequential relationships, such as those between earlier and later titles of a serial, are found in over 73% of relationship records, followed by whole-part relationships for 14.02% of the records. For materials in the music and visual MARC formats, the whole-part relationship predominates, with 56% of the music relationship records and 52% of the visual materials displaying whole-part relationships. Accompanying relationships closely follow, shown in 37% of the music records and 48% of the visual materials records. The latter display no equivalence or sequential relationships. Music materials exhibit no sequential relationships and virtually no equivalence relationships (only 0.03%). For maps, a striking 82% show whole-part relationships, reflecting the ex-

tensive use of contents notes and series statements in map records. Another 18% of the map records exhibit derivative relationships, but less than 1% of map records exhibit equivalence or accompanying relationships; none exhibit sequential relationships.

Another interesting finding pertains to accompanying relationships. We have already seen that accompanying relationships occur most often among visual materials. The accompanying relationships are expressed through the physical description statement in the bibliographic record. Such accompanying information is given a special MARC code, the subfield code "e" within the "300" field for all formats. Table 4a shows the percentage of records in the LC database with subfield "e" as compared to table 4b, which shows the percentage of records in the relationship file. As we can see, accompanying materials, as indicated through the physical description statement, are more prevalent for visual formats than for other formats in the LC database as a whole; essentially the same

TABLE 4a

PERCENTAGE OF LC DATABASE RECORDS WITH FIELD 300 SUBFIELD "e"

Format	No. of Records with Subfield "e"	No. of Records in this Format	% of all LC Records
Books	1,838	2,319,079	0.079
Serials	16	290,312	0.006
Maps	46	87,416	0.053
Visual	33,629	79,335	42.389
Music	362	20,940	1.729

TABLE 4b

PERCENTAGE OF RELATIONSHIP FILE RECORDS WITH FIELD 300 SUBFIELD "e"

Format	No. of Records with Subfield "e"	No. of Records in this Format	% of Relationship Records
Books	1,846	2,325,515	0.079
Serials	16	320,634	0.005
Maps	265	101,408	0.026
Visual	32,313	79,275	40.761
Music	366	21,760	1.682

percentage is found in the relationship file.

Some other accompanying relationships can be found expressed in the MARC serials format in two special fields, the "525" note field and the "770" supplement/special issue entry field, which specify accompanying material that is either a supplement or special issue. The results show 1.97% of records in the relationship file for serials are expressed by a "525" note and another 0.52% by a "770" when no "525" was present, for a total of 2.49%.

An interesting finding for sequential relationships is that over 47% of serial records contain "78x" linking fields, other than those for absorbed titles. That is, 151,793 records with specific continuing or preceding links were found in the 1968-July 1986 LC machine-readable database.

SUBJECTS

As noted earlier, subjects were grouped into five categories based on preselected ranges of Library of Congress classification numbers: science, social sciences, humanities, general, and unknown.¹⁷ Unfor-

tunately, during the study it was discovered that the Library of Congress is not consistent in including one and only one 050 field for the Library of Congress classification number; this is particularly true for music and serials, and for some types of visuals.¹⁸ This fact, of course, skewed the data for music and serial records but fortunately did not affect the data collected for visuals, maps, or books.¹⁹ As a consequence this report on findings omits data for serials and music but includes data for books, maps, and visual materials.²⁰ Because serials were omitted, there are no instances of sequential relationships represented. Figure 5, then, shows the distribution of occurrences for each relationship type in each subject area for all items except serial and music items.²¹

There is a highly significant degree of difference among the subject categories with respect to types of bibliographic relationships. This difference can be seen clearly in figure 5, a bar graph of the distribution of subjects for all formats studied. Unfortunately, we do not have comparative data for the LC database as a whole by subject for books, maps, and visual materials.

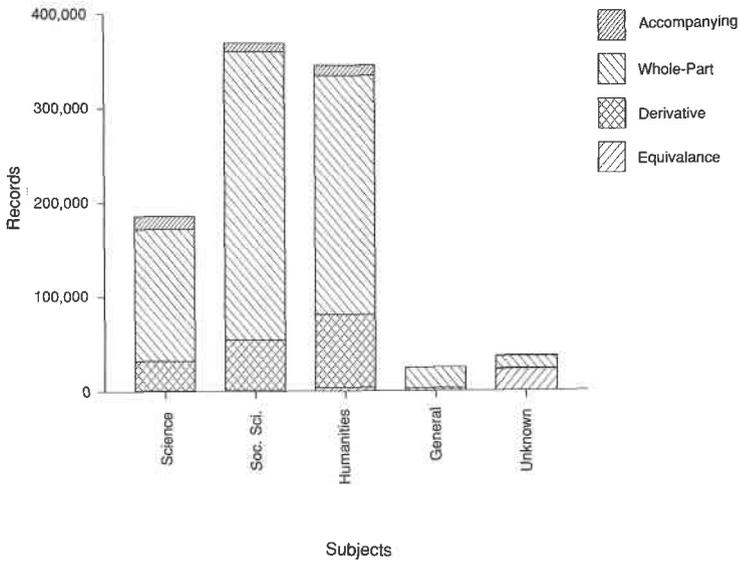


Figure 5. Distribution of Relationship Type by Subject.

The percentage of records in the "unknown" subject area, particularly for equivalence relationships (82.88%), reflects the practice of assigning "MLC" call numbers to microforms rather than LC classification numbers. For all bibliographic records containing relationship information, apart from the "unknown" category, there is the expected similar percentage for the humanities (35.95%) and social sciences (38.37%) typical of a major research library, while the hard sciences constitute only 19.33% of the relationship file.

As for general conclusions from figure 5 regarding types of relationships by subject, it is interesting to see that each category of relationship is associated with its own predominant subject: accompanying relationships occur more often among science materials (39.44%) than among materials in other subjects; whole-part relationships occur most often among social science materials (41.70%); derivative relationships occur most often among materials in the humanities (46.73%); and

equivalence relationships, reflecting microforms without classification numbers, occur most often in the "unknown" subjects (82.88%). Remember that these data omit serials and music. If we examine the distribution of relationship records within the subject categories, whole-part relationships occur most often, followed by derivative, accompanying, and equivalence relationships for every subject except the "unknown" subject category, in which equivalence relationships occur most often. Materials in the "unknown" subject category predominantly exhibit equivalence relationships, followed by whole-part, accompanying, and derivative. Obviously, the LC practice of assigning the "MLC" number, an unknown subject, rather than an LC classification number to microforms distorts the count of equivalence relationships.

The subject data can also be examined in terms of particular MARC formats, as in figure 6. As can be seen, the predominance of a subject category varies by bibliographic format. For books, the major sub-

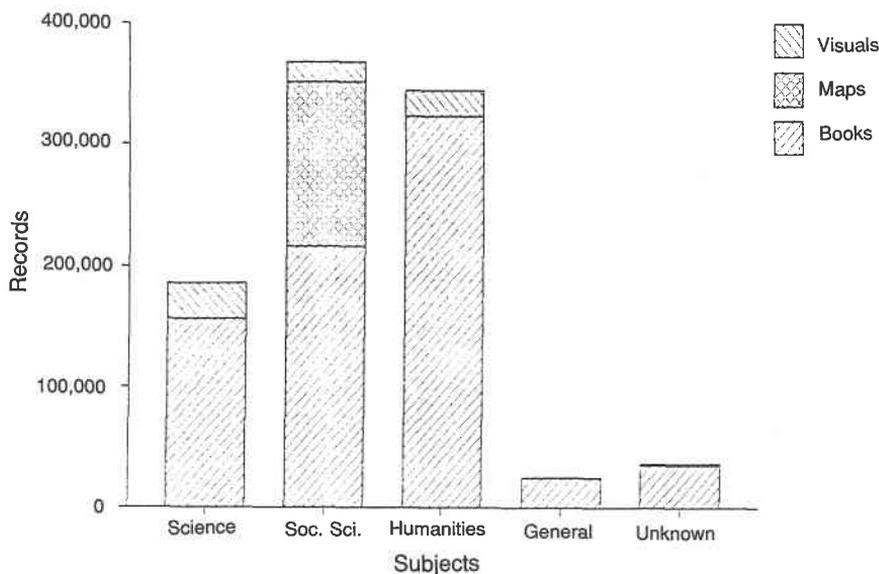


Figure 6. Distribution of Formats by Subject over Relationship Records.

ject area is the humanities (42.83% of all books records) followed by social sciences (28.70%), science (20.73%), and general (3.18%), with "unknown" accounting for 4.55%. For maps, the overwhelming majority were classified in the G (geography) classification, which was categorized as social science (99.51%), with 0.002% in the sciences, none in humanities or general, and 0.488% unknown. Visual materials were a bit more evenly divided among subject areas, with 42.46% in the sciences, 31.02% in the humanities, 24.36% in social sciences, and 0.77% in general, leaving 1.39% unknown.

PUBLICATION DATES

Publication dates are grouped by pre-1700, 1700s, 1800s, and then by decade to the present. Figure 7 graphically presents the occurrences of relationships by publication date.²² The frequencies with which different relationships occur by publication date are highly significant.

When we examine the differences between periods, we see that nearly 5% of the relationship file are for materials published during the 1800s, 0.43% in the 1700s, and only 0.04% pre-1700. This would be expected because the LC database is very skewed in favor of post-1968 publication dates (materials in the database were cataloged after 1968) and the decade for the 1980s was only a little more than half complete by mid-1986. Most of the items in the database as a whole and in the bibliographic records contained in the relationship file were published during the 1970s.

It is interesting to speculate on other causes of differences. When we look at types of relationships, we see a comparatively high percentage of equivalence relationships in the 1800s. A possible explanation is an ongoing serials microfilming preservation project at the Library of Congress. As for sequential relationships, they are predominant during the 1800s and from 1910 through 1949. We might speculate that

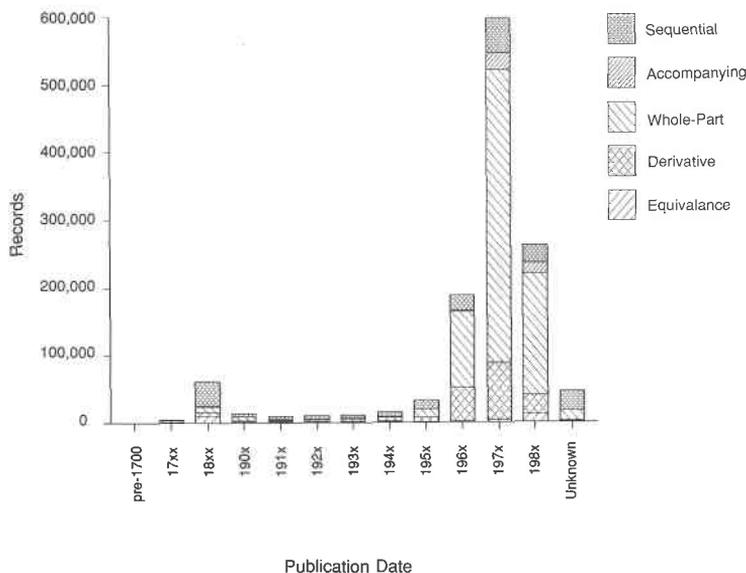


Figure 7. Distribution of Relationship Type by Publication Date.

the CONSER project to catalog serials retrospectively has had an impact on LC's cataloging, so that there is a correspondingly higher percentage of serial records than records for other formats for those early years. But in fact, in the database as a whole, except for the 30,862 serial records in the 1800–1899 period (10.63% of the serials) and 4,755 serial records in the pre-1700s, book records occur more often than serial records.

We would expect each decade to show a pattern of increased level of publication corresponding to the trends in publication in general, with the much-discussed exponential growth of publication since 1952.²³ Because the materials in the database are merely a reflection of the cataloging done at the Library of Congress since 1968, the occurrences of records with pre-1968 dates are skewed by cataloging projects, such as the CONSER project mentioned above. Even so, we can see the publication volume increases gradually from 1910 onward. There is a relatively

high percentage of items published during the 1900–1909 period, reflecting the unusually high percentage of music records (12.84%) for that period. The number of records during the 1900–1909 period in the relationship file is exceeded only by the periods from 1940–1949 onward. The drop from 1910 through 1939 may be attributed to World War I, the Depression, and the diminished publication patterns for books in the United States as documented in Sterling.²⁴ In fact, during the years from 1910 through 1914, general U.S. publishing was very high, matched again only in the post-1953 period. It is also curious that, when we project the number of records for the relationship file for the 1980s based on the number already cataloged, we find the increase is not as great as that between the 1960s and the 1970s, perhaps again reflecting the starting cataloging date for records included in the LC database, but perhaps also reflecting a decrease in publishing after 1974, also noted by Sterling.²⁵

There is also a category of unknown dates, which accounts for 3.69% of the relationship file. These unknown dates reflect cataloging for items without publication dates and items whose bibliographic records have dates containing major typographical errors.

If we look only at the period from 1960 through the 1980s, we find an increase in whole-part and accompanying relationships with a drop in derivative and sequential relationships. Equivalence relationships have gone down and up. It is hard to find an explanation for these changes. Perhaps there are now more items in series and fewer editions of works than there were in the 1960s, but the changes could be due to internal cataloging priorities at the Library of Congress.

LANGUAGES

As would be expected, English, represented by 65.3% of the relationship records, is by far the predominant language. The next closest language is German, represented by only 7.69%. The observed distribution of languages in the relationship file is very similar to the expected distribution in the LC database as a whole, as shown in table 5.

The predominant languages of items having bibliographic records in the LC database are, in order of highest fre-

quency, English, German, French, Spanish, Italian, Russian, and Portuguese.²⁶ This is also the order for the books subset of the LC database as a whole. The distribution for the relationship file can be seen graphically in figure 8.²⁷

The "Other" language category is interesting to analyze. Within this category are the "blank" or "no language" coded values, as well as all other languages not singled out for the study. In the LC database as a whole, the category of "blank" or "no language" accounts for 71.14% of music and 1.83% of visual records, which is the highest percentage after English for visuals. "No language" signifies the lack of text or words. "Other" languages are associated with only 10.16% of the relationship files in the study, signifying materials in languages other than the seven selected for study. The "other" languages constitute 16.07% of the records containing equivalence relationship information.

We find a highly significant degree of difference among frequencies with which different languages are associated with different relationship types. While whole-part relationships dominate, they do so to varying degrees. As to other relationships, an interesting discrepancy is apparent between the romance languages and the Germanic and Russian languages with respect to the derivative relationship. Editions in and translations into French, Italian,

TABLE 5
LANGUAGE DISTRIBUTION COMPARING OCCURRENCES OF BIBLIOGRAPHIC RELATIONSHIPS TO THE LC DATABASE AS A WHOLE

Language	Bibliographic Relationship Records		Records in LC Database as a Whole	
	No.	%	No.	%
English	819,499	65.30	1,773,937	63.42
German	96,484	7.69	198,770	7.11
French	86,983	6.93	161,525	5.77
Spanish	55,753	4.44	120,705	4.32
Italian	36,477	2.91	65,525	2.34
Russian	16,746	1.33	58,096	2.08
Portuguese	15,570	1.24	52,164	1.86
Other	127,450	10.16	366,360	13.10
Total	1,254,962	100.00	2,797,082	100.00

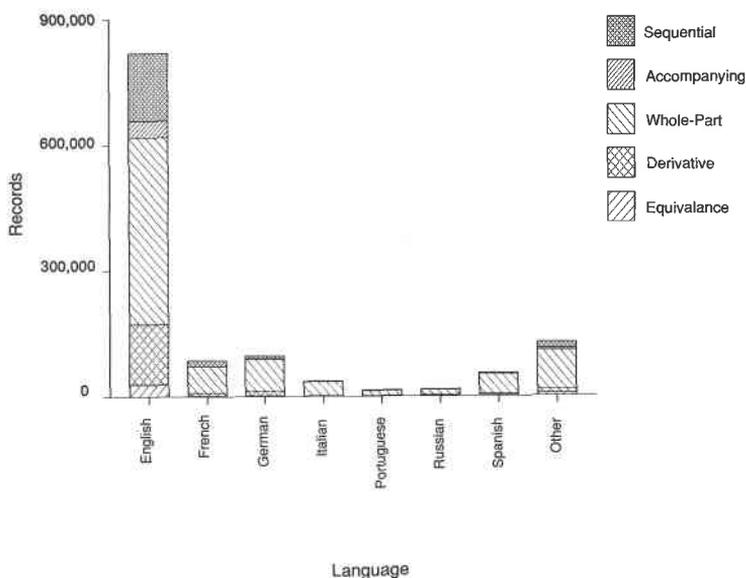


Figure 8. Distribution of Relationship Type by Language.

Portuguese, and Spanish occur much less often than editions in and translations into English, German, and Russian. However, this romance versus Germanic and Russian division does not characterize other types of relationships. For example, equivalence relationships occur comparatively often in records for Russian items (10.18%), but hardly any occur for German (1.46%) or English (3.52%).

Translations and editions deserve closer inspection. Table 6 presents data for derivative relationships, which include translations and editions. Here we find some interesting differences between translations and editions for the various MARC formats. In particular, editions of maps constitute the highest number of derivative relationships for materials in German and Russian. This reflects the fact that Germany and Russia are major producers of cartographic materials along with France, England, and the United States. As for the romance-language items, translations of books constitute the highest

number of derivative relationships for materials in French, Italian, Portuguese, and Spanish.

COUNTRIES OF PUBLICATION

The MARC coded values for country of publication number approximately 300, of which 294 were found on records in this study. In order to simplify the presentation of country of publication, twenty categories of general geographic areas were devised, plus two extra categories, one for undetermined countries and one for typos found in the coding for the records examined. Even the twenty categories are rather unwieldy to present.²⁸

Here again we find a highly significant difference, this time among the country of publication categories associated with bibliographic relationships. As would be expected, the majority of records in the relationship file are for items published in the United States (45.40%), followed by Germany and Austria (7.74%), the United

TABLE 6

DISTRIBUTION OF DERIVATIVE RELATIONSHIPS REPRESENTED BY TRANSLATIONS
AND EDITIONS, BY LANGUAGE
(NUMBER OF RECORDS)

Derivative Relationships	Eng.	Fre.	Ger.	Ita.	Por.	Rus.	Spa.	Other
<i>Translations</i>								
Books	61,469	3,149	3,333	891	315	574	1,644	3,203
Serials	624	30	30	5	4	7	21	95
Maps	0	0	0	0	0	0	0	0
Visual	0	0	0	0	0	0	0	0
Music	177	42	133	116	1	10	9	158
Subtotal	62,270	3,221	3,481	1,012	320	591	1,674	3,456
<i>Editions</i>								
Books	62,509	1,018	1,320	127	107	108	416	1,403
Serials	8,104	1,333	362	40	48	37	212	526
Maps	10,814	1,580	5,872	335	279	1,284	1,032	2,544
Visual	567	2	0	0	0	0	6	7
Music	82	6	54	35	1	3	10	543
Subtotal	82,076	3,939	7,608	537	435	1,432	1,676	4,023
Total	144,346	7,160	11,089	1,549	755	2,023	3,350	7,479

Total records with translation and edition information: 177,751

Total records in derivative relationships: 179,149

Kingdom (7.61%), and Canada and Other Asia (both 4.68%). It is interesting to note that these countries are also included in the 1980 ranking of countries by number of titles published found in Kurian's *The New Book of World Rankings*, which lists the top ten countries as:

1. United States
2. Soviet Union
3. West Germany
4. United Kingdom
5. Japan
6. France
7. Spain
8. South Korea
9. China
10. Canada²⁹

The rankings found in the present study are seen in figure 9. We might expect country of publication to correlate with language, and although our data were not

collected in a fashion to correlate language and country statistically, we can see similarities in the number of occurrences for the different relationships, with one notable exception, the U.S.S.R. For equivalence relationships, the percentage of records in the country category for the U.S.S.R. is 5.94%, which is considerably lower than the percentage of records in the category for the Russian language, 10.18%. Perhaps this reflects a thriving business for translation into Russian, or possibly it is the policy at LC to acquire such translations regardless of where they are published.

It must be remembered that in this portion of the study only those fields within the LC-MARC records that explicitly coded bibliographic relationships were examined. We now turn to the field with general information, namely the "500" field for general notes.

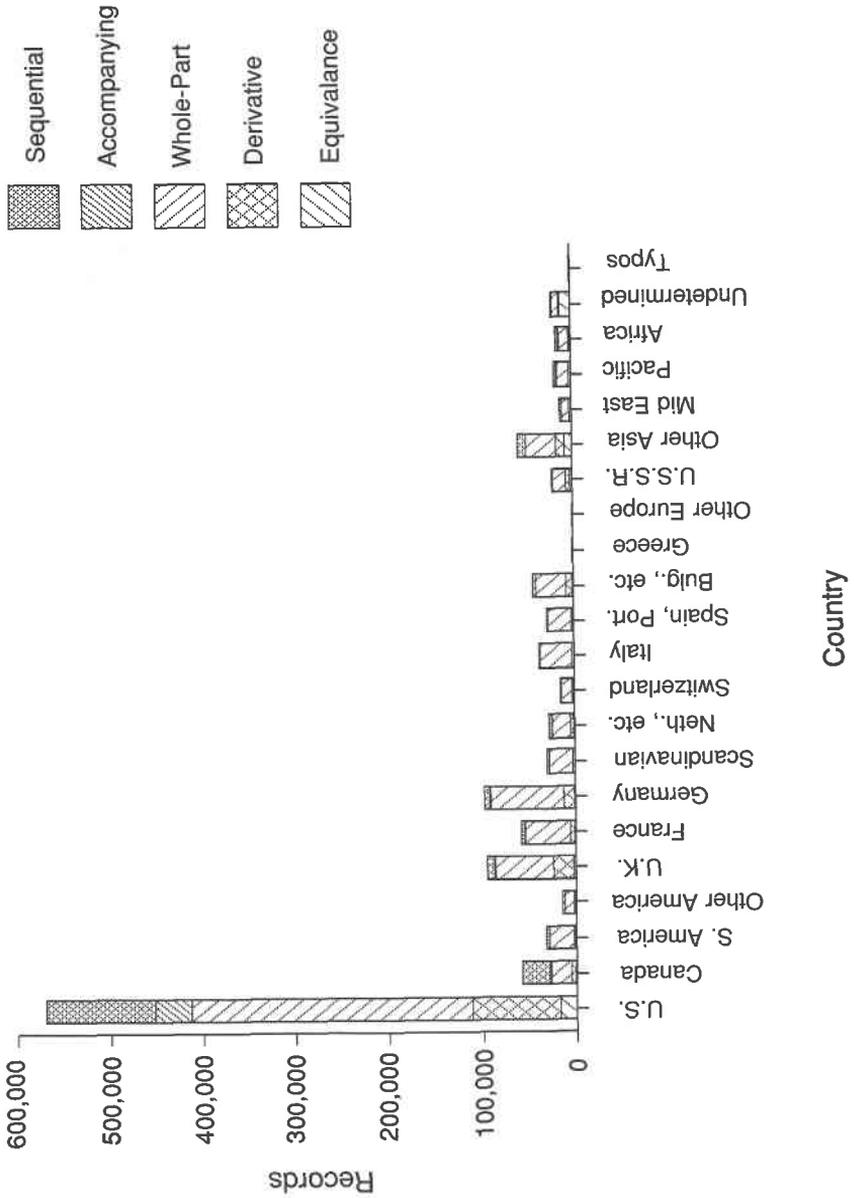


Figure 9. Distribution of Relationship Type by Country of Publication.

STUDY OF "500," GENERAL NOTES: METHODOLOGY

A sample was needed for studying relationships exemplified by the "500" field for general notes, as previously explained under "Counting the Frequency of Occurrence of Relationships." Because general notes contain both relationship and nonrelationship information, each note had to be scanned to segregate those notes expressing relationships. Fortunately, the scanning was simplified by producing a list of keywords within the notes. The keyword list was then scanned by the researcher for relevant terms, so that only those notes corresponding to relevant terms were reviewed.

The sample size needed for a 97% confidence level with a standard error of 3% was determined to be 1,344. The formula for sampling proportions, when the population size is large, is:

$$n = t^2 \frac{pq}{V}$$

where $V = d^2$ and $q = 1 - p$

for n = sample size

d = margin of error (.03)

$V = 0.0009$, using the formula above

p = proportion (using the most conservative value set at 50% or .5)

$q = .5$ (using the formula above)

t = number of standard deviations from the mean, which is 2.2 for a 97% confidence level.³⁰

A reportedly random sample of 1,841 records selected from the LC database as part of a preservation study was opportunely made available through Robert M. Hayes. The selected records had been retrieved from the database of the OCLC Online Computer Library Center, Inc., in the OCLC-MARC display format. Shirley Nordhaus of the UCLA Office of Academic Computing wrote a brief program to reduce the OCLC records to just those fields needed for the study.³¹

Subsequently, the *Oxford Concordance Program* was used to list keywords occurring within the note fields that pertain to relationships and to count their frequency of occurrence. An alphabetical

word list, or index to the concordance, was also generated through the *Oxford Concordance Program*. The word list consisted of all keywords arranged with their frequency of occurrence. The concordance itself was arranged alphabetically under a given frequency of occurrence. A sample page from the wordlist is shown in figure 10, and a sample page from the General Notes Concordance produced through the *Oxford Concordance Program* is shown in figure 11. The results of findings from this study of "500," general notes, were compared to the findings from the study of explicitly coded relationships using the chi-square test, which is the principal statistical test for nominal data. The discussion of findings follows.

STUDY OF "500," GENERAL NOTES: DISCUSSION OF FINDINGS

Normally, notes on bibliographic records clarify information given in the descriptive portion of the bibliographic record or add significant bibliographic information that the cataloger deems useful. Some of the information pertains to bibliographic relationships. Even information that normally would have its own field, such as a contents note with its "505" tag, is found in the general notes studied. In fact, the results indicate that every category of bibliographic relationship, except the shared characteristic relationship, is represented in the general notes. The predominant relationship type is the accompanying relationship, with 326 occurrences, which is found on 17.708% of the sample of 1,841 records. Derivative relationships closely follow with 302 occurrences, then equivalence with 123, whole-part with 40, sequential with 5, and descriptive with 2.³²

No new relationship categories were discovered in the examination of the "500" notes. It was originally presumed that rare relationships might be embedded within the general notes, but no such relationships were discovered by this researcher within this sample. This finding, significantly different from what was expected, indicates that the taxonomic categories derived from the analytical study appear to

GENERAL NOTES CONCORDANCE					
Abdullah 1				Abdullah - af	
Abridgment 1	given by Sheikh Abdullah was first "Nine illustrations" Cover title. Abridgment of thesis (PH	3104274 m 770127 eng i i i 1974 DS			
Abstracts 1	notes on slipcase. Abstracts also in French	7318205 m 810408 eng mdu 1936 E4			
according 1	six heretical teachers according to Series title	506625 m 721205 eng ### 1965 QE			
acknowledge 1	of papers which acknowledge support from Description	700530 m 730808 eng ii 1972 BQ			
Ad 1	an etching by Ad. Laiauze after Nolin	6403703 s 800610 eng mdu a 1973 DC			
Addendum 1	of United Theological "Addendum" and "Errata": C2]	745054 m 731120 eng ### 1876 PQ			
Additional 1	Blank pages for "Additional recipes" at end	6914595 m 801020 eng xx i 1980 DS			
Additions 1	expressed in her "Additions and corrections": C4]	3832266 m 780427 eng pau 1929 TX			
address 1	text of an address made . . . before the	1071224 m 741111 slo pau 1944 PG			
Adrianus 1	The author considers Adrianus at Epictitus to	3926244 m 780525 eng nyu 1945 DS			
advocating	was mainly political, advocating peace and the	2208388 m 750503 ger gw 1955 M1			
Aeronautics 1	Virginia Division of Aeronautics, <Feb. 1961>: Kept	4965980 s 790515 eng nyu a 1813 cr			
af 1	P.: Festschrift udg. af Kobenhavns universitet /	6751202 s 800925 eng vau 19uu TL			
		1218722 m 850621 dan dk 1915 B1			

be exhaustive. However, because the study was based on a sample, it may be that other relationships do in fact exist; but, given this study, they would have to be very rare indeed.

If we generalize using the percentages from the sample of "500," general notes, the number of occurrences of each type of relationship within the population can be estimated. In fact, from the distribution of records in the sample, we find a 97.5% confidence level with a standard error of 0.025, using the formula described above.

Table 7 presents the predicted occurrences for each relationship type within the entire population. This table shows that 43% of the sample contain relationship information in the "500" notes; it is also interesting to observe 70% of the records with "500" notes are records with relationship information.

If we compare the distribution of types of relationships in the relationship file identified by explicitly coded MARC fields with the distribution in the relationship records identified within the general notes

TABLE 7
PROJECTED NUMBER OF RECORDS OF EACH RELATIONSHIP TYPE FOUND IN "500,"
GENERAL NOTES³³

Relationship Type	Occurrences in the Sample		Projected No. of Records with Rel. Information in "500" Notes
	No. of Records	% of Sample	
Equivalence	123	6.681	92,208
Derivative	302	16.404	226,401
Descriptive	2	0.109	1,504
Whole-Part	40	2.173	29,991
Accompanying	326	17.708	244,398
Sequential	5	0.272	3,754
Total	798	43.347	598,256

Sample size = 1,841

Records in the sample with "500" fields = 1,141

Total population size = 2,797,082

Records in the population with "500" fields* = 1,380,157

TABLE 8
DISTRIBUTION OF EXPLICITLY CODED RELATIONSHIPS COMPARED TO
RELATIONSHIPS OCCURRING IN "500," GENERAL NOTES³⁴

Relationship Type	Occurrences in "500" Note Records Sample		Occurrences in the Uniquely Coded MARC Field Records	
	No. of Records	% of Column	No. of Records	% of Column
Equivalence	123	15.41	42,000	3.35
Derivative	302	37.84	179,000	14.27
Descriptive	2	0.25	0	0.00
Whole-Part	40	5.01	782,000	62.36
Accompanying	326	40.85	49,000	3.91
Sequential	5	0.63	202,000	16.11
Total	798	99.99	1,254,000	100.00

Critical chi-square = 14.4494 for df = 6 and alpha = 0.025

Observed chi-square = 7.105

field, as in table 8, we find the distributions vary. The distributions show that sequential and whole-part relationships are more likely shown by explicitly coded fields, while equivalence, derivative, descriptive, and accompanying relationship information are more likely to be embedded in a general note. The disparity between the distribution of relationship information in uniquely coded MARC fields and in "500," general note fields, is confirmed in the chi-square test, which indicates a highly significant degree of difference.

SUMMARY OF EMPIRICAL DATA ON BIBLIOGRAPHIC RELATIONSHIPS

Most records containing bibliographic relationships can be classified as records for English-language items published in the United States between 1970 and 1979. An exception is material involved in equivalence relationships, which was predominantly published in the 1980s. For books, maps, and visual materials, the only formats with subject data, science is the primary subject associated with accompanying relationships, social science with whole-part relationships, humanities with derivative relationships, and the unknown subject category with equivalence relationships. A majority of equivalence relationship records have an "MLC" classification, which was considered an unknown subject. The format with the highest number of records for equivalence, derivative, and whole-part relationships is books, while the formats exhibiting accompanying or sequential relationships with the highest number of records are visual materials and serials, respectively. The number of map records for whole-part and derivative relationships is much higher than that expected from the distribution of maps found in the LC database as a whole.

From the data collected in the study of unique MARC codes, we find the following factors characterize records containing information for each type of relationship:

- Records displaying *equivalence relationships*: 63.34% of them are for books (this is 78% of what would be expected if formats were distributed similarly in the relationship file and the LC data-

base as a whole), 82.88% are in the "unknown" subject category, 13.05% are in the humanities, 68.88% are in the English language, 40.46% are published in the United States, and 30.44% are published in the 1980s.

- *Derivative relationship* records: 79.03% of them are for books (this is about 3% fewer than the expected percentage of records for books found in the LC database as a whole), 46.73% are in the humanities, 80.74% are in the English language, 52.41% are published in the United States, and 46.56% are published in the 1970s.
- *Descriptive relationship* records: as characterized by the two incidences found in the "500" general notes study, these are for books in the humanities in the English language, published in the United States; one was published in 1975 and the other in 1980.
- *Whole-part relationship* records: 75.01% of them are for books (this is 8% fewer than the expected percentage of records found for books in the LC database as a whole), 41.70% are in the social sciences, 56.99% are in the English language, 38.72% are published in the United States, and 55.37% are published in the 1970s.
- *Accompanying relationship* records: 65.98% of them are for visual materials (this is an astonishing twenty-four times higher than the expected percentage of records for visual materials found in the LC database as a whole), 39.44% are in the sciences, 81.30% are in the English language, 79.68% are published in the United States, and 49.66% are published in the 1970s.
- *Sequential relationship* records: when we add the data from the study of "500" general notes, we find 99.46% of the records with sequential relationship information are for serials (this is 880% greater than the serials distribution within the LC database as a whole). From the study of the explicit MARC codes we find 79.42% of sequential relationship records are for items in the English language, 57.74% are published in the United States,

and 25.34% are published in the 1970s. No subject data are available for sequential relationships. Another 0.54% of the records with sequential relationship information are records for books, as found in the study of "500," general notes.

- *Shared characteristic relationships* were not studied.

Basically, we have information about the characteristics of bibliographic items exhibiting relationships and information about the MARC format with respect to relationships. These data tell us that factors of language, country of publication, and publication date characterizing items involved in bibliographic relationships (i.e., in the relationship file) in general closely resemble the same factors characterizing items in the LC database as a whole. However, for the factor of MARC formats there is a dissimilarity. As noted above there is in the relationship file a much higher number of visual records for accompanying relationships, map records for derivative and whole-part relationships, and serial records for sequential relationships than in the LC database. We can then say accompanying relationships will occur most often among visual materials; sequential relationships among serials; and whole-part, derivative, and equivalence relationships among books. On the other hand, if we examine the possibility of predicting the type of relationship from the format, we find books, maps, visual materials, and music items show more whole-part relationships than any other (79.24%, 82.10%, 51.55%, and 55.67%, respectively), while serials show sequential relationships most frequently (73.35%). If we predict by language, materials in all languages most often show whole-part relationships (54.35% of the English-language materials, 72.67% of the French, 79.08% of the German, 89.18% of the Italian, 80.82% of the Portuguese, 68.54% of the Russian, and 84.32% of the Spanish). Likewise, by country of publication whole-part relationships occur most often, except for publications from Canada, which exhibit sequential relationships most often. And for publication date, whole-part relationships occur most often, except for pre-

1700 dates, where derivative relationships occur most often (45.74% of all relationship types), and for 1800-1899 and 1910-1949, where sequential relationships occur most often.

To get an idea of the total number of records exhibiting bibliographic relationships in the 1968-July 1986 LC MARC database, the number of records counted in the study of explicitly coded relationships can be added to the number of records found in the study of "500," general notes. Even then, we do not have a complete count, because we have taken into account neither those bibliographic relationships for which no MARC code exists nor shared characteristics relationships. However, we do know from the data that the total database of 2,854,252 records has a projected 598,256 records containing relationship information in the general note fields and 1,254,000 records have relationship information in the fields designated by explicit MARC codes. To these we can add the unanalyzed 291,000 book records having "440" series statements (whole-part relationships). This gives an estimated total of 2,143,256 records containing relationship information, which is 75% of the total database. However, it must be remembered that the method used to count frequency of records leads to some unknown percentage of records being duplicated in the various types of relationships, so the actual number of records containing relationship information is less than 75% of the total database.

Regarding the number of records containing relationship information, there are several interesting findings. Of 1,254,000 records containing bibliographic relationship information in the study of explicitly coded relationships, 42,000, or 3.35%, are for equivalence relationships; 179,000, or 14.27%, for derivative relationships; 782,000, or 62.36%, for whole-part relationships; 49,000, or 3.91%, for accompanying relationships; and 202,000, or 16.11%, for sequential relationships. A future study should be conducted to determine the cross-correlation of the factors for a more precise estimate of the percentage of a database reflecting bibliographic relationships.

CONCLUSION

Present bibliographic records include a degree of redundancy of bibliographic information. This redundancy takes the form of tracings. If we assume that the future bibliographic record format should optimize the capabilities of computer software, we could reduce the redundancy now evident in the MARC format. For example, filing and display requirements could be embedded in the record more elegantly than the current practice of repeating information to be used as access points. As for bibliographic relationships within the bibliographic record, the important factors are (1) an identifying citation for or other link to the related work and (2) explicit coding or other mechanism to identify the type of relationship. The exact devices for storage and transmission of bibliographic relationships information in the MARC formats or future machine-readable bibliographic formats are topics for research.

Now that we have reached the next technological step in the evolution of the catalog with the computerized catalog, cataloging rules and mechanics of recording need to progress accordingly to take advantage of the new technology. Obviously, the reason all of this is so interesting now is that online catalogs and computer systems provide us with the opportunity of linking records and tagging information to facilitate construction of displays and printed products. Just as Panizzi intended to make the catalog of optimum utility to his library users within the bounds of time and money constraints, we should be thinking of ways to optimize the computer's capabilities. This, of course, is a continuing thread that connects all catalogs and their corresponding catalog rules. We have reached another milestone in technology, just as we did at the turn of the century with the advent of printed cards. And just as the sharing of printed catalog cards changed the cataloging rules and the structure of the catalog entries, the computer-based catalog will certainly make its impact on the evolution of the catalog entry and bibliographic structure.

REFERENCES AND NOTES

1. It should be noted that the unauthenticated CONSER records were not included in this study because they were not included in LC's serial file.
2. Moreover, the Library of Congress creates only one bibliographic record for each item cataloged. One bibliographic record for each item cataloged means there is one record for each separately published item, and there may be one record for each part of a set as well as a record for the set as a whole. Therefore, using the LC database avoids the problem of skewed data caused by duplicate records, which are common in some other online databases.
It is also interesting to note that the Library of Congress database is probably the only file reflecting the original cataloging of a single library. The databases for most other libraries include records created by the Library of Congress and contributing members of the national bibliographic databases.
3. The figures are from OCLC's "Bibliographic Records in the Online Union Catalog by Source of Cataloging, 1986 October," in the March 12, 1987, OCLC Pacific Network *News Update*, no. 25.
4. Also included under relationships with specific codes are access point relationships that involve a shared characteristic, such as shared author. The Library of Congress has counts of the frequency of occurrence of name headings and subject headings in all files as reported in 1981 by Sally McCallum and James L. Godwin.
At the time of this study the Library of Congress had no software to select a true random sample from its databases. In February 1987 they acquired such software.
5. Table 9 (on next page) was derived from "Descriptive Tabulation, Library of Congress MUMS Format Data" for each format studied.
6. Another possibility, of course, is a relationship not specifically included in previous bibliographic records. This latter possibility was excluded from this study after a fruitless search for other bibliographic relationships than those from the taxonomy during the study of general notes. If extremely rare relationships do exist, they are left to a future study.
7. See Barbara B. Tillett, "A Taxonomy of Bibliographic Relationships," *Library Resources & Technical Services* 35:150-58 (1991).
8. The table of targeted tags, indicators, subfield codes, and values can be found as

TABLE 9
RECORDS WITH FIELDS TAGGED "500" IN THE LC DATABASE

Date of Tabulation	Format	Total Records	No. of Records with Tag 500	% of Total Records
6/29/86	Books	2,319,079	1,069,422	
1/17/86	Serials	290,312	146,233	
1/24/85	Maps	87,416	84,988	
7/17/86	Visuals	79,335	60,515	
5/20/86	Music	20,940	18,999	
	Total	2,797,082	1,380,157	49.34

appendix B in Barbara Ann Barnett Tillett, *Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Relationships Used in Cataloging* (Ph.D. diss., Univ. of California, Los Angeles, 1987), p.303.

9. Counting such relationships would require reviewing all access points that two or more records might have in common.
10. In two cases there were fewer than five records, which were counted as no records.
11. Every few years the Library of Congress counts tags, indicators, and subfield codes as part of its database maintenance through the program "Descriptive Tabulation, Library of Congress MUMS format data" created by James L. Godwin, Automated Systems Office, Library of Congress. The printouts from the most recent "Descriptive Tabulation" were used to verify whether any records existed for a particular query.

The Library of Congress has two additional computer programs for statistical analysis, written by James L. Godwin, called *JANUS Bibliographic Retrieval System*, and its subprogram, *JLGSTATS*, to extract frequency data on given tags, fields, subfields, and text strings within fields.

The bibliographic format became the primary division for the computer runs, because the Library of Congress has its database in several files according to bibliographic format. The only files available as of 1986 were books, serials, music (Library of Congress does not distinguish between sound recordings and scores), visual materials, and maps. The archives-manuscripts and machine-readable data files formats were not then available at LC. The name and subject authorities formats were not examined, because only biblio-

graphic relationships were in scope for this study. However, basic counts from the two authorities files were collected for future studies.

Once a specific bibliographic format file was retrieved, a computer analysis was conducted based on the MARC content designators, sometimes to the byte value, for each relationship field as described in the text above. The selection of appropriate MARC fields was made by the researcher using the most current description of the MARC formats and relying on familiarity with the various codes and values that would include bibliographic relationships. After identifying the specific fields, bytes, and values to examine for each of the five MARC formats, Leo Settler coded the queries, with the technical advice of John James, and the coding was then keyed and the programs run by Todd Daniel. All three people worked for the Library of Congress: Leo Settler was from Bibliographic Products and Services, John James from the Automation Planning and Liaison Office, and Todd Daniel from the Cataloging Distribution Service. Not all of the computer runs ran successfully the first time. The Books runs for edition statements ran on less than 20% of the books file. An attempt was made to rerun that set of records, but this was not completed in time for this study.

Another problem came with two runs for books in series. Initially the runs were not received and a rerun was requested. A rerun was received for all but the books' "440" field, traced series titles.

The results of the 119 queries comprise nineteen large three-ring binders.

12. Two other factors were also examined but later dropped from the study, namely:
 6. date of cataloging and
 7. cataloging rules used.

These two factors were initially studied to see if there was an effect of the change of cataloging rules and rule interpretations on the incidence and nature of bibliographic relationships. However, due to the relative newness of the database, all of the constituent bibliographic records were cataloged after the introduction of the 1967 *Anglo-American Cataloging Rules*, and these two factors could not provide the desired information. The resulting data merely indicated the predominance of the current rules.

13. These ratios, when other than one-to-one, are indicated in appendix B (p.303) of the author's dissertation under the column heading "Ratio of Occurrences to Records."
14. A fourth condition, the mixture of bibliographic information within the general note fields, has already been noted above as a complication for the general study and addressed in a separate study described here under "Study of '500,' General Notes."
15. A random sample of the database was taken in April 1987 to study records with this field, but the results are not included in this study.
16. This information is based on the LC statistics for MUMS format data, "Descriptive Tabulation, Library of Congress MUMS format data" volumes for "MUMS BOOKSM, 06/29/86," "MARC SERIALS WORKFILE, 01/07/86," "MUMS MAPS, 01/24/85," "MUMS VISUAL MATERIALS, 07/17/86," and "MUMS MUSIC, 05/20/86."
17. The fifth category, "unknown," was added as a result of the computer runs to indicate the lack of a specific LC class number. A surprising 23.63% of the fields retrieved were from records without an LC class number. Some of these "unknown" subject items were microforms, which are assigned a call number with the prefix "MLC"; some were prefixed "PAR," which indicates a record in process; and some were minimal-level cataloging records with the prefix "MLC," all without LC classification. An attempt was made to clarify why such a high percentage lacked an LC class. In November 1986, John James, of the Library of Congress Automation Planning and Liaison Office, explained that there were some "broken records" which lacked 050s and some records with multiple 050s where the first 050 contains the original call number and the second 050 contains the microfilm replace-

ment, in which case the second number will be the reel number and not a class number. He further explained in February 1987 that there are always multiple 050s in Music for sound recordings, because the first 050 with first indicator 1 is the suggested classification number and the second 050 with first indicator 0 is the label name and number. The same is true for visual materials with two 050s. For motion pictures, multiple manifestations will each have separate 050s but may be included on the same bibliographic record; however, none of those are yet in the LC MARC file. For maps, multiple 050s are illegal, but there could be some in-process records or some records from the old batch file that are not yet cleaned up, although there should not be any. LC's MARC distribution statistics show the mean occurrences of 050 fields per record to be 0.98 for books, 0.51 for serials, 1.00 for maps, 1.00 for visuals, and 1.39 for music, and mean occurrences in records having one or more fields with this tag as 1.00 for books, 1.00 for serials, 1.00 for maps, 1.03 for visuals, and 1.40 for music.

18. According to "Descriptive Tabulation, Library of Congress Internal MARC Format Data, MARC Serials Workfile, Tag 050, 01/07/86," the classification number field, 050, occurs only in 51% of the serial records, a fact unknown to the researcher at the start of the study, but 050s have 1.0 mean occurrences in records having one or more fields with this tag.

"Descriptive Tabulation, Library of Congress MUMS Format Data, MUMS Music, Tag 050, 05/20/86" shows 1.39 mean occurrences per record for music, or 1.40 mean occurrences in records having one or more fields with this tag.

"Descriptive Tabulation, Library of Congress MUMS Format Data, MUMS BooksM, Tag 050, 06/29/86" shows books have 0.98 mean occurrences per record and 1.0 mean occurrences in records having one or more fields with this tag.

For maps, "Descriptive Tabulation, Library of Congress MUMS Format Data, MUMS Maps, Tag 050, 01/24/85" shows 1.0 mean occurrences both per record and in records having one or more fields with this tag.

Visual materials have 1.03 mean occurrences in records having one or more fields with this tag, according to "Descriptive Tabulation, Library of Congress MUMS Format Data, MUMS Visual Materials, Tag 050, 07/17/86."

19. Only one Visuals run for series not traced (490 0) indicated five too many 050 fields out of a total of 760 records retrieved.
20. As indicated in note 17 above, for bibliographic records for sound recordings in the music file, there are always two 050 fields: one with the first indicator "1" to identify the suggested classification number and the other with first indicator "2" to identify the label name and number. For serials the problem is the lack of an 050 for 49% of the serial records. This omission of 050s can be attributed to the fact that the serials file contains in-process records, which are incomplete. It was too costly to rerun all of the music and serial records to adjust for this problem during the period of this study.
21. A table of the number of records in the distribution of relationship type by subject (omitting serials and music) can be found in the author's dissertation as table 5, p.150.
22. A table showing the number of records in the distribution of relationship type by publication date can be found in the author's dissertation as table 6, p.158.
23. *Encyclopedia of Library and Information Science*, v.1 (New York: Dekker, 1968), p.239.
24. Christopher H. Sterling and Timothy R. Haight, *The Mass Media: Aspen Institute Guide to Communication Industry Trends* (New York: Praeger, 1978), p.8-9.
25. Sterling, *The Mass Media*, p.9.
26. The language counts for the LC database as a whole are based on data found under the language field in the above-mentioned "Descriptive Tabulations" for each of the MARC formats.
27. A table of the number of records in the distribution of relationships by language can be found in the author's dissertation as table 8, p.168.
28. A table of the number of records for the distribution of relationships by country of publication can be found in the author's dissertation as table 10, p.172-74.
29. George Thomas Kurian, *The New Book of World Rankings*, 3d ed., updated by James Marti (New York: Facts On File, 1984), p.411.
30. William G. Cochran, *Sampling Techniques*, 3d ed. (New York: Wiley, 1977), p.75.
31. Those fields were record-entry date, description code, language code, publication date, country of publication code, the first two elements of the LC class number, OCLC record number, and "500" field data.
32. A list of keywords and their frequency of occurrence in the sample of General Note Fields arranged by taxonomy category appears in the author's dissertation, in note 139, p.269-74.
33. Further information about the records with fields tagged "500" in the LC database can be found in the author's dissertation as table 13, p.259.
34. The critical chi-square values are from table 8 of the *Biometrika Tables for Statisticians*, v.1, 3d ed., ed. E. S. Pearson and H. O. Hartley (London: Biometrika Trust, 1976), as reproduced in Richard J. Shavelson, *Statistical Reasoning for the Behavioral Sciences* (Boston: Allyn and Bacon, 1981), p.644.

An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Schema for the USMARC Formats

Gregory H. Leazer

An analysis of data elements in the United States Machine-Readable Cataloging (USMARC) formats for bibliographic, authority, and holdings data is presented. Criteria for the examination of USMARC, derived from the theory of the conceptual schema for databases, are (a) types of entities and data elements that are included, (b) the function of each data element, and (c) the rules of form that govern each data element in the USMARC formats. The major finding of the analysis is that there is a high degree of redundancy present in USMARC, and a more rigorous conceptual plan for USMARC is desirable.

One response of the library and information community to the post-World War II information explosion has been the application of computer solutions to maintain and extend bibliographic control over library collections. The history of librarianship is in part a history of the methods and technology of control over the growing bibliographic universe.¹ The invention and improvement of new technologies require the continual assessment and modification of tools for bibliographic control. Tools such as descriptive cataloging codes and classification schedules are under continual revision. This evolution in control techniques includes the develop-

ment of computer systems for bibliographic databases.

The application of computer solutions to problems in bibliographic control is made possible in part by developments in database management systems. Bibliographic database systems improve with new mechanisms such as better user interfaces for online public access catalogs (OPACs) and the development of classification and superthesauri for subject retrieval.^{2,3} The record formats that contain information on bibliographic items must also continue to improve and develop.

Contemporary concepts of database management should be applied to bibliographic

GREGORY H. LEAZER is a doctoral student at the School of Library Service, Columbia University, New York. The author wishes to acknowledge the help and advice of Professor Richard P. Smiraglia. Manuscript submitted October 8, 1991; revised December 2, 1991; accepted for publication January 6, 1992.

databases. Of particular importance for large-scale bibliographic databases is the selection and implementation of an appropriate database model. Little critical thought exists on the topic of selecting a suitable database model among various alternatives. A clear understanding of the objectives and functions of library catalogs and bibliographic databases is crucial for this work. In order to determine the applicability of individual database models at a future date, a critical evaluation of current bibliographic databases should be conducted. An evaluation should lead to further understanding of how bibliographic databases are maintained.

To this end, an examination of the standard record formats for bibliographic databases was conducted. The United States Machine-Readable Cataloging (USMARC) formats for bibliographic descriptions, authority data, and holdings information were examined with the criteria provided by the theory of the conceptual schema, a tool developed for the implementation and management of databases in general.⁴

This paper begins with a discussion of database management in general, focusing especially on the conceptual schema. The development of machine-readable cataloging in the United States is discussed, and the USMARC record as the object of examination is introduced. The research questions are developed as the result of the application of conceptual schema theory to the USMARC formats, and the methodology is explained. A report of the findings and a discussion follow.

BACKGROUND

CONCEPTUAL SCHEMA

"One problem inherent in modeling any subset of the real world is the difference between the human's perception of the enterprise and the computer's need to organize the structures in a particular way for efficient storage and performance."⁵ This leads to three different levels of modeling a portion of the real world: the user's level, the computer's physical level, and a conceptual level that translates one level to the other (see figure 1). The conceptual level provides for "the mapping from the logical

to the physical, or internal, level, describing the semantics of the entities and relationships, including descriptions of connections and consistency constraints."⁶

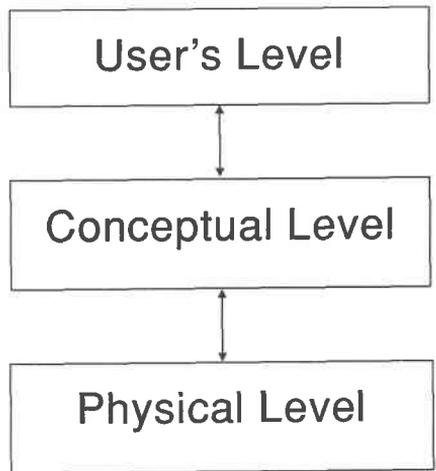


Figure 1. Three Levels of Modeling.

The conceptual level is expressed in the conceptual schema.⁷ The conceptual schema defines the total content of the database.⁸ It describes the complete enterprise of the database—that is, how the database operates and how the data are used.⁹ The schema details what real-world entities are to be included in the database by specifying a "representation of that part of the world that the database is about."¹⁰ It also mandates that decisions on the conceptual level of the database should be made *before* other decisions, such as those concerning the ways that data should be stored, processed, or displayed.¹¹ The conceptual schema also specifies the data elements that are to be recorded on each entity described in the database. "The designer of such a [database] must make decisions about matters such as what elements of information about this material to include. . . ."¹²

A database model is a description of the logical structure of the data contained in a database, that is, of the entities and relationships between the entities described by the model. A database model can

describe the logical structure of the enterprise at any one of the three levels shown in figure 1, although it most commonly does so at the conceptual or physical levels. Harrington states that there are three major models for databases currently in operation: hierarchical, network, and relational models.¹³ Additional models exist, such as semantic data models described by Peckham and Maryanski.¹⁴ Preliminary work toward the application of the relational model to bibliographic databases has been conducted, although without justification for why the relational model appears to be the most appropriate. Selection of the relational model appears to be only a best guess at this time, and alternative database models have yet to be rejected specifically. The application of the relational model falls into at least three categories:

1. The pertinence of relational database models for application to bibliographic databases, especially relational models¹⁵;
2. The application of relational database models for reducing the amount of repetitive information stored in a bibliographic database¹⁶; and
3. The study of relational database models in order to provide a conceptual

framework for the criticism of contemporary cataloging codes and databases.¹⁷

Following the creation of a schema for the database, the appropriate database model is selected. This model will be used to implement the database. Until an assessment of the database is conducted, its structure specified, and the entities, elements, and relationships contained in the database defined, it is premature to select a particular database model.

The conceptual schema also specifies what data are to be collected on each entity described by the database. The database is capable of receiving in one format (an input format), storing in another format, displaying to different groups of users in several display formats, and transmitting to computers in a communications format. While the conceptual schema describes the database independently of any of these various user views, it often directs these formats by specifying which data are required for input or available for output. A format, also called a data schema, is a set of instructions for the formulation and expression of the content of the database. It guides the data creator in the production of the individual records that compose the database (see figure 2). The USMARC

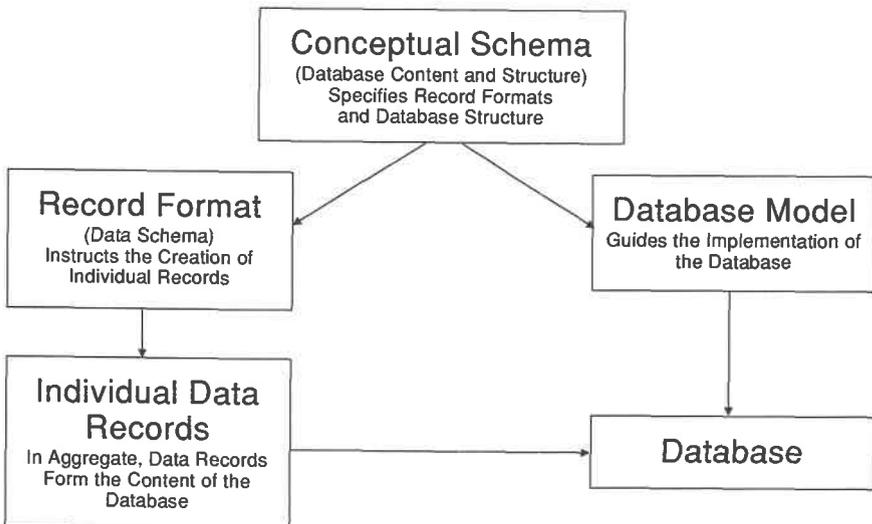


Figure 2.

format is often used for all the different functions of all format types, including input, display, and communication.

No conceptual schema for the USMARC formats is published in the format documents. Neither a formulation of principles nor a description on the conceptual level of the USMARC formats is adequately addressed in the formats themselves. Discussion papers created for the administrating bodies of USMARC could serve as a beginning of a conceptual plan for USMARC. However, they do not fulfill the requirements of a conceptual schema. Information that would be of use in a conceptual schema is provided in the professional literature of library and information science, such as papers by Attig¹⁸ and Weisbrod,¹⁹ but this information must be formally integrated with the USMARC formats. Information for a conceptual schema for the USMARC formats is dispersed through a number of different kinds of documents, including the USMARC formats *per se*, format updates, discussion papers of the committees and agencies that maintain the USMARC formats (such as the MARBI [Machine-Readable Bibliographic Information] Committee of the American Library Association), books, research papers in the professional literature, and the knowledge that resides with the individuals who develop and maintain the USMARC formats. Much of the content and structure of the formats is governed by descriptive cataloging conventions, for example, and is part of the tradition of the library profession, but is not described in any single document. This body of information exceeds the precise definition of the conceptual schema, and perhaps a more rigorous articulation of the conceptual schema is needed.

What is the consequence of the lack of an articulated conceptual schema? At best, the design and evolution of USMARC could be unaffected by the lack of such an articulation. Statements on the objectives and principles of the library catalog, many that predate the computerized catalog (e.g., Cutter's "Objects"²⁰ or Lubetzky's objectives²¹) can stand in lieu of a substantial part of the conceptual schema. Cataloging codes, such as the *Anglo-American*

Cataloguing Rules, 2d edition, 1988 revision (AACR2R),²² can provide definitions for data elements to be included in a bibliographic database, specifying the form of entry for each data element as well. Thus one role of the conceptual schema is fulfilled.

At worst, however, the lack of an articulated schema can result in the partial completion of an effective database system. The lack of a schema could also result in the awkward and unguided evolution of a database. Conventions developed for one form of technology might be detrimental in another technological setting. For example, the card catalog requires a linear file structure so that individual catalog records will file together in order to fulfill the collocation objective of the catalog. With computerized catalogs, linear structures could be replaced by a matrix of connections linking various records together in a nonlinear array. Without a clear statement of the purposes and structure of a database it is impossible to know what database model is appropriate for the bibliographic databases that form the core of today's library catalogs.

How are we to understand the investigations conducted thus far of the application of relational models to bibliographic databases? We can subscribe to the idea of the "expanded" conceptual schema, whereby the schema is dispersed through a number of documents in the professional literature. Thus, newly identified functions for the catalog, such as the formal expression of bibliographic relationships,²³ can become integrated into an expanded conceptual schema. The application of the relational model can more fully express a variety of relationships between individual entities in the database. If we believe that the conceptual schema should be expressed in a single document (which does not currently exist) then it is premature to discuss what kind of database model is appropriate until we have completed such a schema.

Finally, the conceptual schema acts as the foundation for the design of the database. The design of the record format (i.e., the data schema) is derived from the conceptual schema. An inadequate or inconsistent data schema is evidence of

either an absent conceptual schema or a poorly implemented one. Redundancy in the data schema could be intentional to enhance the richness of the data or, perhaps more likely, could be the consequence of poor planning, resulting in higher data-creation and storage costs. Criteria for the evaluation of a record format broadly includes whether it fulfills the requirements of the conceptual schema. In addition, a record format should be compact, efficient, and expressive. Compactness is the economy with which the format records information, its level of redundancy, and its ability to specify information in a concise manner. Efficiency is the economy of human and computer processing effort. Expressiveness relates the record format's ability to express the various characteristics of an entity.²⁴

MACHINE-READABLE CATALOGING

With the advent of computing technology, a machine-readable cataloging format was created in the 1960s. The primary impetus for the creation of computerized bibliographic information was to facilitate the communication of cataloging information created by the Library of Congress (LC). The USMARC format was created "in order to distribute in machine-readable form the same information used to produce catalog cards at LC—and, for most of its first two decades, the primary end-result of USMARC use within libraries was printed catalog cards."²⁵ Machine-readable cataloging was implemented so that catalog cards and other bibliographical products could be printed out easily by the computer.²⁶ Thus the USMARC format was created by the Library of Congress to supplement a card catalog, rather than to replace it. From these inauspicious roots, the USMARC formats grew to become the national standard for the recording, storage, communication, and processing of bibliographic, authorities, and holdings information. Today they continue to be the standard and basis for the national bibliographic utilities, such as the OCLC Online Computer Library Center, Inc., and the Research Libraries Information Network (RLIN). The USMARC for-

ats also form the basis for most local online bibliographic systems, especially in large academic and research libraries. Despite the success of the USMARC formats, there has been little evidence of any major rethinking about the purposes and functions of catalogs since the development stage of the USMARC format.

The USMARC formats form the basis for this study of data elements used in bibliographic databases because they are the national standards for bibliographic, authorities, and holdings data. Although the local implementation of USMARC records varies from setting to setting, USMARC exists as a common record format at the core of almost all major library bibliographic database systems in the United States. Thus the USMARC formats are a valid object of study for the consideration of the design and modification of bibliographic databases. Although many bibliographic database systems that do not use the USMARC formats exist, they lie outside the purview of this study. Such systems include local library catalog systems that do not support USMARC standards and database systems such as *ERIC*, *Library Literature*, *Dissertation Abstracts International*, and others that commonly provide access to the specifically defined, nonmonographic literature of a discipline.

The formats were examined to determine what kinds of entities are described, what data elements are specified, and how individual data elements function. Also, the structure of the record format itself was examined. Because relational database designs are often discussed in the professional literature, relational mechanisms already present in the USMARC formats were noted.

Examination of the data elements in the USMARC formats is a first step toward the construction of a conceptual schema for bibliographic databases. Such an investigation documents machine-readable cataloging as it currently exists. Functions that the catalog of the future might perform are not documented. However, an evaluation of the current format is one aspect of identifying features that might be appropriate for tomorrow's catalog.

Will it be necessary to restructure the

USMARC formats? The answer might possibly be affirmative. Even though the USMARC formats have proven to be remarkably resilient over the years, they might be incapable of further surviving the evolutionary development of bibliographic databases. Further, the reappraisal of data elements in a bibliographic database should result in the compilation of an "extended AACR2, other reformed standards, and the restructured MARC format . . . into a single document dealing with the entire activity of creating machine-readable records."²⁷ As the product of cataloging activity shifts primarily to the creation of machine-readable bibliographic records, cataloging codes will need to be revised to create a total code whose end product is the machine-readable bibliographic record, and record formats must change in order to exploit more fully recent developments in database management.

DEFINITION OF TERMS

There is an inherent difficulty in matching the terms of database management with the technical vocabulary of the USMARC formats. Databases contain information on entities, that is, a database contains representations (usually linguistic) about specific real-world objects. A telephone directory, for example, is a database that describes the entities that are people who have telephones in a defined geographical area. A data element is an individual piece of information about a particular entity represented in the database. Most commonly, entities are described by a number of discrete data elements. In the simple telephone directory database, the people (entities) are described by their names, addresses, and telephone numbers.

The USMARC formats describe the structure of a single USMARC record. Each USMARC record for bibliographic data is a representation of an individual bibliographic entity. A collection of a number of USMARC records makes up a database. A USMARC record itself includes a number of fields, "a string of characters . . . identified by a [numerical] tag."²⁸ Furthermore, an individual USMARC field often comprises a number of subfields. These

subfields can be considered to be equal to a single data element. A field might include one or more specific elements, for example, the Details of Publication field 260 in the format for bibliographic description includes a place, a publisher's name, and a year of publication, each placed in a different subfield.

USMARC, as a record format, is not a conceptual schema. USMARC is a data schema that governs the ways in which data are recorded. This evaluation is in part an investigation of the degree to which USMARC appears to rest upon a firm conceptual foundation.

This investigation is guided in part by the criteria established by the requirements of the conceptual schema. As indicated above, a conceptual plan demands a statement identifying the entities that are represented in the database. In other words, what real-world objects are represented by the database? By definition, a bibliographic database includes information about bibliographic entities.²⁹

There is an inherent difficulty, however, in stating that bibliographic databases record information on bibliographic entities. This is due to a continuing historical confusion about the definition of bibliographic entity. The statement that bibliographic entities are objects collected by libraries (e.g., books, documents, microforms, sound recordings, videotapes, etc.) is overly crude and simple. Lubetzky and Hayes recognized the confusion about the objects that are described by catalogs (and therefore bibliographic databases):

The essence of the modern concept of cataloging has gradually emerged from a growing realization that the *book* (i.e., the material record) and the *work* (i.e., the intellectual product embodied in it) are not coterminous; . . . that the *book* is actually only one representation of a certain *work* which may be found in a given library or system of libraries in different media.³⁰

Today the term *item* is often preferred over the term *book*, specifically because of the variety of nonbook media produced and collected by libraries. To state that bibliographic databases are about bibliographic entities is to ignore this continuing tension in bibliographic control.

Furthermore, there is a tension between the use of the term *entity* in bibliographic control and in database management. Bibliographic entities are described using bibliographic records, which in turn contain a variety of data elements. Many of the specific data elements contained in the USMARC formats could be database entities in and of themselves. For example, a USMARC bibliographic record describes a specific item, but the publisher named as a data element in the bibliographic record would qualify as a database entity. What kind of entity is described by a USMARC authority record? It is not a bibliographic entity; rather it is commonly the name of a person or a subject term. This paper reports an investigation of the presence of potential database entities recorded in the USMARC formats.

Another requirement of a conceptual schema is that it detail those specific data elements that are to be included about each entity. Identification of USMARC data elements would serve as a part of an overall program for the reappraisal of the data elements in bibliographic databases. The conceptual schema should also specify the function of individual data elements. The identification and analysis of data elements will help reveal the function of individual fields in the USMARC formats. The conceptual schema also specifies what standard of control is applied to each data element³¹: Are there general rules for the form of entry or an authority list for each data element? The standard of control for an individual data element forms a continuum from all permissible values enumerated in the USMARC formats to a total absence of any standard where there might be no general rule of form for entry.

USMARC should meet the criteria of a record format. The examination of the USMARC format should determine whether a consistent principle has been applied to the structure of the USMARC formats. Two decades of growth and amendments to the MARC formats might have resulted in a complicated and inconsistent record architecture. Of special concern here is whether USMARC is compact. Repetition of data is evidence of a lack of compactness.

RESEARCH QUESTIONS

Specifically, the following questions guided this study:

1. What kinds of data are included in the USMARC formats? That is, can a typology of data elements be developed?
2. Does the examination of data elements provide evidence of non-bibliographic entities in the USMARC formats?
3. Does the examination of the USMARC formats reveal a pattern among the data elements? Are the data compact?
4. In which fields are certain kinds of data stored?
5. What are the functions of the individual data elements contained in the USMARC formats?
6. What relational mechanisms (if any) already exist in the USMARC formats?
7. Are there general rules of form for each data element? What standard of control is applied to each data element?

This research was conducted in part to support the future design of a more sophisticated bibliographic database. A thorough examination of the formats might reveal a well-ordered structure and demonstrate a consistent application of database principles. This could occur despite the fact that bibliographic databases constructed according to the USMARC formats have not been rigorously described at the conceptual level. Alternatively, structural deficiencies would suggest that a more rigorous conceptual approach is necessary. Further, because the potential database designs include the possibility of relational mechanisms, it is appropriate to determine what relational mechanisms currently exist in USMARC.

RESEARCH METHODOLOGY

A data-gathering device (see appendix A) was created consisting of the numerical tags for each of the fields present in the USMARC formats for bibliographic, authorities, and holdings information. A census of the USMARC formats was then conducted examining each field in turn

and recording its data content, its function, and ways that it related to the overall structure of the format. Appendix B contains examples of completed data-gathering forms.

For example, upon the examination of the first field that dealt with the place of publication, a data-gathering form for place of publication was created, and the field tag number corresponding to the identified field was recorded on it. Upon the subsequent location of another, different field that dealt with the place of publication, that field's tag number was recorded.

The data-gathering forms were then examined for patterns. The degree of conceptual design for the formats was appraised on the basis of the construction and organization of the formats. Organizational problems would indicate the necessity for a rigorous conceptual plan for the USMARC formats.

FINDINGS

Most readily apparent was the high degree of redundancy in the record structures. Although no statistical measurement of redundancy was conducted, there are obvious repetitions of data elements occurring in multiple locations in the formats. Information such as the place and date of publication are recorded in several fields (24 and 58 different fields in the three formats respectively). Similar repetitions were found for several other types of data in the three formats, for example, location of the cataloged item (occurs in 11 separate fields), edition information (18 fields), and form of musical composition (26 fields, not including those fields that express topical content). Table 1 is a list of repetitions in the three USMARC formats.

Even highly specific information is recorded in multiple locations. Frequency of publication is located in 8 separate fields, information on processing and preservation in 6 fields, information on lending and access in 9 fields, and even the expression of the presence of mathematical data is located in 2 fields. Some specific information, such as file characteristics for computer files, is located in only one specific

TABLE 1
FREQUENCY TABLE OF DATA ELEMENTS

Category	No. of Fields
Cataloging Source Information	4
Credits Note on Participant and Performers	2
Date of Publication	58
Date/Time Topical Information	32
Edition	18
Extent	5
File Characteristics	1
Form of Musical Composition	26
Frequency of Publication	8
Geographical Topical Information	36
Lending and Access	9
Location of Item	11
Mathematical Data	2
Physical Size	3
Place of Publication	24
Preservation and Processing Information	6
Price	2
Reproduction Note	6
SEE Information	6
SEE ALSO Information	6
Source Data Information	2
Topical Information	50
Unit of Storage Information	3

field created for that data, not including the catch-all notes fields.

Some redundancies, however, are functional or possibly unavoidable. One major source of redundancy is in the expression of the topical content of the cataloged item. Topical content is expressed in 50 different fields in the three formats. However, some of this redundancy is the result of signifying the same topical concepts in different ways, using different subject access tools. The most familiar of these tools are the *Library of Congress Subject Headings*, the *Library of Congress Classification*, and the *Dewey Decimal Classification*. Analyses of the comparative

advantages of different techniques used for subject retrieval is currently under way³²—whatever the results of this research, a preliminary conclusion must be that this kind of redundancy adds an additional dimension in the expression of subject content and allows for the possibility of more flexible and successful subject retrieval mechanisms.

The measure of redundancy in the USMARC formats is directly related to how restrictively data elements are defined by the researcher. Information on the place of publication, while redundant in the USMARC formats, is merely a subset of all fields dealing with details of publication or of all fields that express geographical information (72 fields in total). The USMARC format currently makes a distinction between personal, corporate, and meeting names: personal names are recorded in 14 fields, corporate names in 16, and meeting names in 12. Broadening the scope to include names of all types, a higher number of fields would be found. The same thing occurs for a number of other fields that currently exist as separate and distinct data elements but have the same or similar data content and as such could conceivably be merged into a single data element in a new bibliographic record schema. Certain types of fields suggest themselves as families of a type of data content, such as title fields (50 individual fields in total, including 22 fields of uniform titles), all the fields that describe physical characteristics (19 fields), and a number of similarly related fields already mentioned, such as fields with geographical or date content. Table 2 is a list of families of data elements of similar content.

The redundancy of data elements spread out over a number of fields is only one indication of the possible lack of a clear conceptual organization for the USMARC formats. Repetitions of the same or similar data (i.e., the element families) usually do not occur together in the USMARC formats. Fields with similar content are scattered all over the formats. Some of the patterns of data-topography in the USMARC formats were intentionally planned, for example, 5XX fields are notes,

TABLE 2
FREQUENCY TABLE OF DATA
ELEMENTS, BROADLY DEFINED

Category	No. of Fields
Added Entries	23
Bibliographic Relationships	26
Corporate Names	16
Date and Time Information	85
Geographic Information	72
Language	38
Main Entries	7
Meeting Names	12
Notes	51
Personal Names	14
Physical Description	19
Series Information	33
Title Information	50
Uniform Titles	22

6XX provide subject access, and X10 are corporate name headings (but not 010, 210, 310, or 510) in the bibliographic format. Also, most of the eight areas specified in the International Standards for Bibliographic Description (ISBD) are translated in their general order in the USMARC formats, with title information and the statement of responsibility occurring in fields 20X–24X, edition information in the 250–29X fields, physical description in 3XX, series information in 4XX, and notes in 5XX. Around the core of the narrative-type ISBD-family fields (20X–5XX) are shaped additional narrative information fields controlling access points for names, subjects, and other added entries. Preceding all of this narrative information is coded information.

An examination of the function of the individual fields was conducted. It was considered useful to identify those fields that fulfilled a goal of the catalog in certain broad categories: (a) to identify bibliographic items, (b) to collocate items, or (c) to evaluate items in the library collection.³³ Unfortunately, it is not easy to identify the specific function of individual fields. While certain groups of fields are established

purely to provide access, e.g., 1XX and 6XX fields in the bibliographic format, it proved difficult to sort out these functions for the majority of fields easily. Fields used primarily for the identification of items, such as the bibliographic Title Page Transcription field 245 and the bibliographic Physical Description field 300, also help the user evaluate or make informed choices among items or works represented in the online catalog. Field 245 commonly generates an added title entry point. In the online catalog, every field and word becomes a potential access point.

Present in the USMARC formats are a number of fields whose function is to identify uniquely the item being cataloged. These "absolute identifiers"³⁴ appear in 67 separate fields in the three USMARC formats, including 12 fields that record more than one absolute identifier. Not included in this count are some fields that are intended to identify items uniquely, for example, call numbers with cutter numbers or personal name headings. These 67 fields include International Standard Serial Numbers (ISSN), which appear in 28 separate fields, International Standard Book Numbers (ISBN), Standard Technical Report Numbers, copyright article fee codes, matrix numbers for sound recordings, universal product codes, etc.

Bibliographic relationships are handled by 26 special fields maintained for this purpose. There are 13 fields in the 76X-79X range in the bibliographic format for linking a parent record to other bibliographic records in the database. In addition to these 13 fields for bibliographic relationships, there are 13 other fields that express these relationships. Depending how one interprets the purpose of some fields, one can determine that they express certain bibliographic relationships. The General Notes 5XX fields are also capable of expressing bibliographic relationships, for example a statement such as "Microfilm reproduction of original published: London : J. Murray, 1859. xvi, 123 p." These notes fields were not counted as specialized fields for bibliographic relationships.

There is another group of specialized fields used for linking records. There are 9

fields in the three formats that connect independent USMARC records together using control numbers, primarily for relating holdings records to their parent bibliographic records. Together with the 13 fields in the 76X-79X range, there are 22 fields that relate specific records to one another. Table 3 is a list of the fields that share a specialized function in the three USMARC formats.

TABLE 3
FREQUENCY TABLE OF IDENTIFIED
FIELDS WITH SPECIALIZED FUNCTIONS

Category	No. of Fields
Absolute Identifiers	67
Bibliographic Relationships	26
Record-Specific Relationships	22

There are two different mechanisms provided in the USMARC formats for linking records. There is an automatic mechanism for linking records that relies upon a mechanism in an online bibliographic system design in order to function. This mechanism automatically relates individual bibliographic records or an authority record to a bibliographic record, for example. The latter example exists in the Western Library Network (WLN) and is described by Culp.³⁵

The linking of entries for bibliographic relationships, however, is achieved primarily through the provision of headings in narrative forms, without any use of an explicit linking mechanism in an online system. These headings function in manual, card-based systems by collocation. The history of specific techniques to express bibliographic relationships is analyzed by Tillett.³⁶ In the computer environment, narrative headings work the same way, but they fail to exploit the ability of the online catalog to connect records automatically. Because mechanisms that automatically relate individual records to one another already exist, the USMARC formats contain fields for the expression of these relationships, and catalogers are already making notes about bibliographic relation-

ships, librarians should be able to create a database that can effectively express these relationships using the capabilities of the computer.

Table 4 is a typology of identified data elements in the USMARC formats. This table is an attempt to define generally what kind of data is to be found in a USMARC database. Category 1, Titles, is self-explanatory. Titles are a common and familiar data element in bibliographic descriptions. Non-title Item Identifiers are the absolute identifiers discussed above, such as ISBN and ISSN. The Names category (category 3) contains the names of both individuals and corporate bodies, as well as variant forms of names that are recorded in the USMARC authorities format. These names can identify a number of different functions in relation to the bibliographic entity, such as authorship or topical subject. Names also serve as identifiers for a person or corporate body that could be contained in a related file of "creating persons and bodies" entities. Date and Geographic Information might contain a similar number of different kinds of relationships to the bibliographic entity as well. Date information, for example, can relate to a date of publication, a chronological subdivision in a topical subject heading, or the date of capture for a recorded performance. Bibliographic Rela-

tionships includes series information (Bibliographic 4XX) and edition statements.³⁷

Category 7, Physical Descriptions, and category 8, Intellectual Descriptions, are less familiar. Specific data elements in Physical Descriptions include pagination for books or duration for sound recordings, for example. Intellectual Descriptions can include the form and genre of the item being described, the medium of performance for music scores, topical information, and the presence of bibliographies, indexes, or mathematical data.

Category 9, Local Information, includes all information that is specific to a particular library concerning the item being described. Such data elements might include information on lending or access, circulation, preservation or processing, or holdings data. Category 10, Explanatory Notes and References for Cataloging Practice, includes information on the establishment of headings, including Source Data Found and Source Data Not Found notes and scope notes for the application of headings.

There are structural similarities between different categories as well. For example, titles and non-title identifiers both serve to name or specify individual bibliographic items or works. The categories presented are not mutually exclusive, as a Name (category 3) could also be present as an Intellectual Characteristic (category 8)—for example, in a topical subject heading for a biography. Also, Geographic Information is usually expressed by a name but could also be described by geographic coordinates. The categories in table 4 do not necessarily exhaust all data elements present in the three USMARC formats, either. Information such as frequency of publication for serials is either not included in the categories as presented or has a tentative existence either as an Intellectual or a Physical Characteristic. Furthermore, some individual data elements do not belong to one specific type of entity. Titles, for example, are strictly a data element of bibliographic entities; date information, however, could relate to the date of birth of an author (person entity type) or the date of copyright for a book (bibliographic entity type).

TABLE 4
 TYPOLOGY OF DATA ELEMENTS
 PRESENT IN THE USMARC FORMATS

-
1. Titles
 2. Non-title Item Identifiers
 3. Names
 4. Date Information
 5. Geographic Information
 6. Bibliographic Relationships
 7. Physical Descriptions and Characteristics
 8. Intellectual Descriptions and Characteristics
 9. Local Information
 10. Explanatory Notes and References for Cataloging Practice
-

It was established above that the conceptual schema of a database design should not only identify all of the entities to be included in the database but also define the specific data elements to be recorded about each entity. The conceptual schema should specify the values for each data element that are valid. This is achieved either by using an authority list of all possible values of the data element or by specifying their general rules of form. At the highest level of standardization, there are 61 different fields among the three USMARC formats that have small, fixed authority control such that the individual cataloger cannot change the domain of valid entries. The domain for these fields is set and defined in the USMARC format documents. These fields are all exclusively in the leader, the variable control fields (00X), and the variable data fields (01X-09X) of the three USMARC formats. Although the valid entries sometimes are mnemonic, they often are not. The principle of different storage formats could be extended here, where a system design integrated with a machine-readable cataloging format could store the data in a coded form and display the information in narrative form.

Many fields exist where there is a slightly lower level of standardization and control for the form of the data content but authority control still exists. These fields are of the kind for 1XX and 6XX entries, where the form of the entry is controlled by an external database of authorized entries, such as the Library of Congress Name Authority File (LCNAF) or the *Library of Congress Subject Headings*. Here, however, the individual cataloger is given more leeway in the creation of non-authorized headings. A number of different authority control mechanisms exist for the maintenance of standardization, and there has been discussion about increasing the number of fields under this kind of authority control.³⁸ Fields of this type are prime candidates for relational structures of the type utilized by WLN mentioned above.³⁹ For example, when assigning a personal name to a bibliographic record, a cataloger first checks local files and the LCNAF for the authorized form of the

name. After failing to locate an authorized form of the name, the construction of personal name headings is controlled by AACR2R. Inclusion of this name in a USMARC record is then controlled by the rules for individual fields, in this case, field 100.

Many other fields exist at a lower standard of control whereby only the form of the entry is controlled. An exhaustive list of possible values is not maintained. These fields are controlled by a combination of rules for entry provided by AACR2R and the USMARC formats. Rules for locating data in a bibliographic item and for transcribing and formatting it are used together with field content designation, indicator values, and subfield codes in the creation of bibliographic records.

Finally, there are fields where there is practically no standard for data content. The 5XX Notes fields in the bibliographic format, especially the 500 General Note field, are the best examples of this standard. Some of the fields specify what kind of information should go in them, for example, the 502 Dissertation Note field, but have little or no specification of general rules of form beyond the examples provided for the field or explanation in AACR2R. The General Note 500 field is a collection of all note information that will not fit into any other note field, and there is no information beyond the examples in the USMARC format about form of data content; all information on data content is derived from the descriptive catalog codes and the unwritten practices of catalogers. There is no apparent principle at work explaining why some kinds of specific information receive a special notes field—for example, information on Numbering Peculiarities gets a special note (field 515 in the bibliographic format)—but others do not, including the very common indication of the presence of indexes.

ANALYSIS

The examination of the USMARC formats reveals a very complex record structure in which a large number of discrete data elements are recorded. It is unclear whether all the data contained by USMARC are

attributes of bibliographic entities. A way of perceiving data elements that share a common, broadly defined characteristic, such as those listed in table 2, is to view them as separate entity types, distinct from documents or other bibliographic entities. Certainly not all of the entity types in table 2 would qualify as nonbibliographic entities (for example, titles and added entries are data elements of bibliographic entities). However, geographic places, languages, dates, and the names of people and corporate bodies could be attributes of distinct nonbibliographic entities that have been mixed into the USMARC formats for bibliographic data. A relational database could keep different entity types in separate files and then relate the entity types to bibliographic items. This would allow the relationship between entities to detail the function that one entity serves in relation to another. A file of geographic entities, for example, could be related to bibliographic entities in the following ways:

document <published in> place,
 document <is about> place,
 document <stored in> place.

Furthermore, a file of geographic entities could have relationships within it between the individual geographic places, such as:
 place1 <is in> place2.⁴⁰

Possible candidates for entity types currently included in a bibliographic database defined by the USMARC formats include bibliographic entities, people, corporate bodies, places, dates, and languages. Each of these entity types generally embraces a large number of individual data elements and USMARC fields.

The most obvious kind of redundancy is the number of fields that express highly similar data, although perhaps in different forms. The existence of 24 separate fields for data on place of publication is an example of this. While no individual USMARC record will include all 24 fields for place of publication, one could wonder whether all 24 fields are necessary. Examination of the redundant fields present in USMARC begins to reveal *repetitions* of data—the same data might be included in several places in a USMARC record. A precise measurement of these repetitions would require an examination of actual

USMARC records, but this study has begun to demonstrate these data repetitions.

Two patterns of data repetition are readily apparent. One pattern is the repetition of narrative data following the expression of the same conceptual information in coded form. The mathematical data redundancy is a simple example: the presence of mathematical data is expressed in two fields in the bibliographic format, the 034 Coded Mathematical Data field and the 255 Mathematical Data Area field. The same information is signified in two different ways to two different ends. Field 034 is used for coded information to best exploit the storage and sorting capabilities of the computer; field 255 is used to convey the same information in narrative form for human, rather than computer, consumption.

This kind of repetition, however, serves no practical purpose. Here, the ability of the computer to store data in one format and display it in another format can be exploited. The data in field 034 would be best suited for computer storage, and a computer algorithm could translate those data into a narrative form upon input or display, reducing the duplication of effort a librarian must spend in cataloging the item. If the data were translated upon display, then this would effectively reduce the amount of space required to store this information in the computer's memory. This pattern of repetition, once for the computer and once for the human, occurs in several places, for example, with data on place of publication, date of publication, intellectual content of the item, and information on its physical description.

Further examination of the repetitive data phenomenon reveals an additional pattern of redundancy. The USMARC bibliographic format contains the 76X-79X linking entry fields, where similar data elements are recorded in order to provide complete bibliographic information for bibliographic relationships. The data contents of the 76X-79X fields are not necessarily repetitions of the same information contained elsewhere in the bibliographic record; rather, the 76X-79X information is data repeated from related bibliographic records. A better mechanism for linking related records might be to employ the

relational abilities of the computer, automatically linking records without repeating data from one record into another.

Another kind of redundancy uncovered derives not from the repetition of the same data, but rather from the multiple recording of information to fulfill a single function of the bibliographic record. Certain functions of the catalog have traditionally required several discrete data elements; for example, bibliographic items are uniquely identified in part by the recording of physical details. These details include several data elements, especially dimensions, pagination, and the presence of illustrations for books. As mentioned above, the USMARC formats seem to require an overwhelming amount of data. It would be useful to evaluate critically all data elements to discern those elements that are really necessary and those that are included because of the weight of tradition. Hufford recently studied the user behavior of reference librarians to discover the portions of the bibliographic record that are actually used. He found that "reference staff members generally consulted a limited number of the elements in the bibliographic records."⁴¹ To include excessive amounts of data that are not exploited by catalog users is wasteful, and one method of economizing would be to eliminate data elements that are not needed to fulfill a function of the catalog or are redundant.

An example of this kind of data overkill is the recording of absolute identifiers for bibliographic items. Sixty-seven fields are given over to absolute identifiers. These identifiers have the ability to identify uniquely and relate items in the bibliographic database with other nonlibrary databases, such as the warehouse stock databases of book vendors. Although such features are not now widely acknowledged as functions of the library catalog, the inclusion of such data will help create more flexible and volatile databases. However, 67 fields of this information might be too much. Consolidating these absolute identifiers and reducing their number are beyond the resources of librarians alone, but the need to reduce the amount of redundancy in the USMARC formats includes

reducing the number of identifiers recorded. Sixty-seven fields is not an insignificant portion of the record format structure itself, and important absolute identifiers should be identified, along with the specific function of the catalog each fulfills.

An examination of the USMARC formats reveals a plethora of individual fields of different data elements and of different types. The USMARC format documents, by intention, are an agglomeration of specifications with very little articulation of cataloging principles, either for the cataloging of bibliographic items or for expressing data in machine-readable form. The translation required to take an AACR2R record and turn it into a USMARC record involves a step up in complexity: data must be mapped out, indicators and subfields coded, and non-narrative coding information (much of it a repetition of information included in the manual catalog record) must be created.

There is little evidence of the use of a conceptual schema for the design and maintenance of the USMARC formats. The formats themselves contain only a limited amount of description on the conceptual level. Furthermore, this investigation of the USMARC fields—the data elements included and their identified functions—did not reveal evidence of a clearly articulated and comprehensive conceptual plan for the formats. The redundancy, structural deficiency, partial inclusion of nonbibliographic entities, confusion of functions of individual data elements, and lack of articulation of the rules of entry for some data elements do not demonstrate the presence of a clear conceptual plan for the formats. On the contrary, this research provides evidence of a format in need of a more rigorous conceptual design.

CONCLUSION

A major finding of this study is the large degree of redundancy in the USMARC formats, which is evidence of the lack of a successfully implemented conceptual schema. In some instances, such as the expression of subject content, this redundancy

can serve to increase the effectiveness of the database. In many cases, however, the redundancy is a pure repetition of data present elsewhere in the format. This kind of redundancy is uneconomical and leads to unnecessary complexity. The data redundancy and complexity of the formats demonstrate that substantial conceptual work on the purposes and objectives of machine-readable cataloging is still needed. The redundancy and complexity of the formats also represent a return of the repressed after a successful paring down and simplification of descriptive rules for cataloging.

A crisis in cataloging was declared in 1941 after the rules for cataloging became increasingly complex and a decline in catalog production resulted in large arrears of uncataloged items.⁴² A general project, caused at least in part by this crisis, was carried out in order to simplify descriptive cataloging practices. Lubetzky proceeded through the descriptive cataloging code rule by rule and asked whether each individual rule was necessary. Lubetzky asked, "Are all these rules necessary? Are all the complexities inevitable? Is there an underlying design that gives our code unity and purpose?"⁴³ This quotation translates remarkably well for application to the USMARC formats by rephrasing it: "Is this field necessary? Is this data element inevitable? Is there a conceptual schema that gives our bibliographic databases and record formats unity and purpose?" Again, today, the Library of Congress is faced with a massive rearrange of 38 million uncataloged items.⁴⁴ This new crisis provides cause for the reappraisal of machine-readable cataloging practices and a formulation of the conceptual principles behind the USMARC formats.

Recent research and development on USMARC has focused on the problem of integrating the seven USMARC material-specific bibliographic formats into a single format.⁴⁵ In 1980, MARBI and the Library of Congress began to bring the various descriptive formats together, and with a study conducted by Weisbrod in 1981, format integration became a recognized goal of MARBI and the MARC advisory group.⁴⁶ A principle liability of the various

descriptive USMARC formats was their difficulty in handling complicated materials such as nontextual serials and items of mixed media.⁴⁷ An additional benefit of format integration was maintaining consistency in the USMARC formats, thus lowering costs of maintaining the single integrated format.⁴⁸

The process of format integration made four different types of changes: extensions, making a data element that was valid in one particular format valid for all materials; obsolesces, where specific data elements that were not useful are no longer available in new records; deletes, removing a data element from the USMARC formats if the designator had been reserved but had never been used; and adds, which meant that a new data element was added to USMARC.⁴⁹

The process of format integration was guided by investigating the USMARC formats for those data elements that are necessary (the process of extending or adding data elements) or those that are present but no longer necessary (the process of deleting or making obsolete certain elements). One of the goals of format integration was to "weed out useless elements rather than integrate them to additional forms of material."⁵⁰ It is unclear what specific criteria were used in the determination of whether a data element was useless or not, but the majority of changes were extensions. A very few elements were added, only a limited number of elements were deleted (six fields and two subfields), and 16 whole fields and 23 subfields were made obsolete.⁵¹

While format integration did not use an articulated conceptual schema for the evaluation of the seven descriptive formats, format integration was the result of a conceptual decision that the number of varying formats for bibliographic description should be reduced. Format integration will result in the use of a single format that will be more complicated than any one of the previous seven formats because the number of adds and extensions over the number of deletes and obsolesces resulted in a net increase in the number of data elements in the new single format. The new format itself will be more complicated

than any one descriptive format prior to integration.

The development of a conceptual plan for bibliographic databases should be based upon an understanding of the user's needs. Following an analysis of user needs, a clear conceptual schema for the overall organization and structure of bibliographic databases should be created. The conceptual schema of the current USMARC formats is inadequate, and the structure is confused. This has resulted in a large number of redundancies. There are no expressed principles for the inclusion, placement, functions, or rules of form for most USMARC fields. The conceptual schema also should include justifications for each entity and relationship included in the database. For each data entity and relationship, specifications for their inclusion should be articulated and their organization specified. In addition, the design of a new conceptual schema should not be limited merely to the entities and data elements already present in the USMARC formats, but should recognize that new technology could allow librarians to include data not currently recorded.

That the USMARC formats have been found to be lacking in structure and to contain redundant information should come as no surprise to those people who work with the formats. According to Svenonius, "full-level cataloging, particularly as rendered in the MARC bibliographic formats, is probably wasteful and excessive; it is certainly redundant."⁵² The results of this study supply evidence in support of such statements.

REFERENCES AND NOTES

1. For example, see Ruth French Strout, "The Development of the Catalog and Cataloging Codes," *Library Quarterly* 26:254-75 (1956). Strout traces the development of bibliographic control technology from the earliest library artifacts to Cutter's catalog and code.
2. See, for example, Arlene G. Taylor, "Authority Control and System Design," in Sheila S. Intner and Richard P. Smiraglia, eds., *Policy and Practice in Bibliographic Control of Nonbook Media* (Chicago: American Library Assn., 1987), p.76.
3. See, for example, Nancy J. Williamson, "The Role of Classification in Online Systems," *Cataloging & Classification Quarterly* 10, no.1/2:95-104 (1989), or Karen Markey Drabenstott and others, "Analysis of a Bibliographic Database Enhanced with a Library Classification," *Library Resources & Technical Services* 34:179-89 (1990). See, for example, Marcia Bates, "Rethinking Subject Cataloging in the Online Environment," *Library Resources & Technical Services* 33:400-412 (1989).
4. *USMARC Format for Bibliographic Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1988-); *USMARC Format for Authority Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1987-); *USMARC Format for Holdings Data: Including Guidelines for Content Designation* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1989-).
5. Joan Peckham and Fred Maryanski, "Semantic Data Models," *ACM Computing Surveys* 20:154 (1988).
6. *Ibid.*
7. Dagobert Soergel, *Organizing Information: Principles of Data Base and Retrieval Systems* (Orlando, Fla.: Academic, 1985), p.137-64; and M. J. R. Shave, "Entities, Functions, and Binary Relationships: Steps to a Conceptual Schema," *Computer Journal* 24:42-47 (1981).
8. Soergel, *Organizing Information*, p.138.
9. C. J. Date, *An Introduction to Database Systems*, 4th ed. (Reading, Mass.: Addison-Wesley, 1986), p.36.
10. Raya Fidel and Michael Crandall, "The AACR2 as a Design Schema for Bibliographic Databases," *Library Quarterly* 58:125-26 (1988).
11. Shave, "Entities, Functions and Binary Relationships," p.42.
12. Fidel and Crandall, "The AACR2 as a Design Schema," p.125.
13. Jan L. Harrington, *Relational Database Management for Computers: Design and Implementation* (New York: Holt, 1988), p.50.
14. Peckham and Maryanski, "Semantic Data Models," p.153-89.
15. Robert G. Crawford, H. Sue Becker, and Janet E. Ogilvie, "A Relational Bibliographic Database," *The Canadian Journal of Information Science* 9:21-28 (1984). Clifford Alan Lynch, *Extending Relational*

- Database Management Systems for Information Retrieval Applications* (Ph.D. diss., Univ. of California, Berkeley, 1987); Clifford A. Lynch, "Nonmaterialized Relations and the Support of Information Retrieval Applications by Relational Database Systems," *Journal of the American Society for Information Science* 42:389-96 (1991).
16. Terrence A. Brooks, "Repetitive Bibliographical Information in Relational Databases," *Journal of the American Society for Information Science* 39:2-7 (1988).
 17. Fidel and Crandall, "The AACR2 as a Design Schema," p.123-42.
 18. John Attig, "The USMARC Formats: Underlying Principles," reprinted in Crawford, *MARC for Library Use*, 2d ed. (Boston: Hall, 1989), p.307-25; and John Attig, "The Concept of a MARC Format," *Information Technology and Libraries* 2:7-17 (1983).
 19. David L. Weisbrod, "NUC Reporting and MARC Redistribution: Their Functional Confluence and Its Implication for a Re-definition of the MARC Format," *Journal of Library Automation* 10:227-39 (1977).
 20. Charles A. Cutter, *Rules for a Dictionary Catalog*, 4th ed. (Washington, D.C.: Govt. Print. Off., 1904).
 21. Seymour Lubetzky, *Cataloging Rules and Principles: A Critique of the A.L.A. Rules for Entry and a Proposed Design for Their Revision* (Washington, D.C.: Processing Department, Library of Congress, 1953).
 22. *Anglo-American Cataloguing Rules*, 2d ed., 1988 rev., prepared by the Joint Steering Committee for the Revision of AACR, ed. Michael Gorman and Paul W. Winkler (Chicago: American Library Assn., 1988).
 23. Barbara B. Tillett, "A Taxonomy of Bibliographic Relationships," *Library Resources & Technical Services* 35:150-58 (1991).
 24. Soergel, *Organizing Information*, p.150-52.
 25. Crawford, *MARC for Library Use*, p.171.
 26. Henriette D. Avram, *MARC, Its History and Implications* (Washington, D.C.: Library of Congress, 1975), p.3.
 27. Michael Gorman, "New Rules for New Systems," *American Libraries* 13:241-42 (1982).
 28. Crawford, *MARC for Library Use*, p.333.
 29. Soergel, *Organizing Information*, p.207.
 30. Seymour Lubetzky and R. M. Hayes, "Bibliographic Dimensions in Information Control," *American Documentation* 20:248 (1969).
 31. Soergel, *Organizing Information*, p.142.
 32. Research on the topic of using classification for subject retrieval was recently reported in Songqiao Liu and Elaine Svenonius, "DORS: DDC Online Retrieval System," *Library Resources & Technical Services* 35:359-75 (1991).
 33. The three functions of the catalog are explained in Richard P. Smiraglia, "Bibliographic Control Theory and Nonbook Materials," in *Policy and Practice in Bibliographic Control of Nonbook Media* (Chicago: American Library Assn., 1987), p.17-18.
 34. This is the term used by David Bearman in "Authority Control Issues and Prospects," *American Archivist* 52:290 (1989).
 35. Gwen Miles Culp, "Authority Control within the Washington Library Network Computer System," in Mary W. Chikas, ed., *Authority Control, the Key to Tomorrow's Catalog: Proceedings of the 1979 Library and Information Technology Association Institutes* (Phoenix, Ariz.: Oryx, 1982), p.62-84.
 36. Barbara B. Tillett, "A Summary of the Treatment of Bibliographic Relationships in Cataloging Rules," *Library Resources & Technical Services* 35:393-405 (1991).
 37. Edition and series information is treated as a kind of bibliographic relationship in Ronald Hagler and Peter Simmons, *The Bibliographic Record and Information Technology* (Chicago: American Library Assn., 1982), p.74-78.
 38. For example, see David Bearman, "Authority Control Issues and Prospects," and David Bearman and Peter Sigmond, "Explorations of Form of Material Authority Files by Dutch Archivists," *American Archivist* 50:249-53 (1987).
 39. See Culp, "Authority Control within the Washington Library Network Computer System," also, Fumiko H. Coyne and Ingrid Mifflin, "Shared Authority Control at the Western Library Network," *Library Resources & Technical Services* 34:494 (1990).
 40. This model is described in Bearman, "Authority Control," p.292.
 41. Jon R. Hufford, "Elements of the Bibliographic Record Used by Reference Staff Members at Three ARL Academic Libraries," *College & Research Libraries* 52:54-64 (1991).
 42. Andrew D. Osborn, "The Crisis in Cataloging," in Michael Carpenter and Elaine Svenonius, eds., *Foundations of Cataloging: A Sourcebook* (Littleton, Colo.: Libraries Unlimited, 1985), p.90-103.
 43. Seymour Lubetzky, *Cataloging Rules and Principles: A Critique of the A.L.A. Rules*

- for Entry and a Proposed Design for Their Revision (Washington, D.C.: Processing Department, Library of Congress, 1953), p.1. Emphasis mine.
44. "Library Develops Plan to Cut Arrearages," *Library of Congress Information Bulletin* 49:170 (1990).
45. Elaine Svenonius, "Bibliographical Control," in *Academic Libraries: Research Perspectives*, ed. Mary Jo Lynch (ACRL Publications in Librarianship, no.47) (Chicago: American Library Assn., 1990), p.45.
46. At this time there are different formats in USMARC for bibliographic description: Archival and Manuscripts Control, Books, Computer Files, Maps, Music, Serials, and Visual Materials.
47. Crawford, *MARC for Library Use*, p.224.
48. Stephen P. Davis, "Format Integration: Handling Serials and Mixed Media," *Information Technology and Libraries* 9:162-67 (1990).
49. Sally H. McCallum, "Format Integration: Handling the Additions and Subtractions," *Information Technology and Libraries* 9:156-57 (1990).
50. McCallum, "Format Integration," p.155.
51. *Ibid.*, p.156-60.
52. Crawford, *MARC for Library Use*, p.221-22.

APPENDIX A
DATA-GATHERING FORM

Bibliographic				Authority				Holdings	
				500			Leader/05	100	Leader/05
Leader/05				501			Leader/06	110	Leader/06
Leader/06				502		700	Leader/17	111	Leader/17
Leader/07				503		710	008/06	130	004
Leader/17				504	550	711	008/07	150	007
Leader/18	040	100		505	555	730	008/09	151	008/06
Leader/19	041	110		506	556	740	008/10	260	008/07
	042	111	300	507	561	752	008/11	360	008/08-11
007	043	130	302	508	562	753	008/12	400	008/12
	044	210	306	510	565	754	008/13	410	008/13-15
008/06	045	211	308	511	567	755	008/14	411	008/16
008/07-10	046	212	310	512	570	760	008/15	430	008/17-19
008/11-14	047	214	315	513	580	762	008/16	450	008/20
008/15-17	048	222	320	515	581	765	008/17	451	008/21
008/18-34	050	240	321	516	582	767	008/28	500	008/22-24
008/35-37	051	242	330	518	583	770	008/29	510	008/25
008/38	052	243	331	520	584	772	008/32	511	008/26-31
008/39	055	245	340	521	585	773	008/33	530	010
	060	246	350	522	59X	775	008/38	550	014
010	061	247	351	523		776	008/39	551	020
011	070	250	362	524		777	010	663	022
015	071	254		525	600	780	014	664	023
017	072	255		527	610	785	020	665	024
018	074	256		530	611	787	022	666	027
020	080	257		533	630		040	640	030
022	082	260	400	534	650		042	641	035
023	086	263	410	535	651	800	043	642	583
024	088	265	411	536	653	810	045	643	841
025	09X		440	537	655	811	050	644	843
027			490	538	656	830	052	645	845

APPENDIX A CONTINUED

Bibliographic				Authority	Holdings	
028	540	657	850	053	646	852
030	541	69X	851	060	667	853
032	544		880	070	670	854
033	545		886	072	675	855
034	546			073	678	863
036	547			082	680	864
037				083	681	865
				09X	682	866
						867
						868

APPENDIX B
SELECTED DATA FORMS

Edition Information		Date/Time Topical Information		
Bibliographic Format		Bibliographic Format	Authority Format	
250	772	045	Leader/05	400 *y
534	773	523	Leader/06	410 *y
600 *s	775	600 *y	008/17	411 *y
700 *s	776	610 *y	045	430 *y
760	777	611 *y	100 *y	450 *y
762	780	630 *y	110 *y	451 *y
765	785	650 *y	111 *y	500 *y
767	787	651 *y	130 *y	510 *y
770	800 *s		150 *y	511 *y
		655 *y	151 *y	530 *y
		656 *y		550 *y
		657 *y		551 *y

Place of Publication			Topical Information			
Bibliographic Format		Authority Format	Bibliographic Format		Authority Format	
008/15-17	760 #d	643	050	600	008/07	09X
044	762 #d		051	610	008/15	150
257	765 #d		055	611	008/17	151
260 #a	767 #d		060	630	043	260
260 #e	770 #d		061	650	045	360
265	772 #d		070	651	050	450
518	773 #d		071	653	052	451
533 #b	775 #d		072	655	053	550
534 #c	776 #d		080	656	060	551
550	777 #d		082	657	070	646
	780 #d		09X	69X	072	680
	785 #d			754	073	681
	787 #d				082	
					083	

Fields with Date or Time Content

	Bibliographic Format			Authority Format			Holdings Format
008/06	300 #a	600 #d	700 #d	800 #d	045	500 #d	008/08-11
008/7-10	310 #b	600 #f	700 #f	800 #f	100 #d	500 #y	008/26-31
008/11-14	315 #b	600 #y	710 #d	810 #d	100 #y	510 #d	
033	321 #b	610 #d	710 #f	810 #f	110 #d	510 #y	
045	362	610 #f	711 #d	811 #d	110 #y	511 #d	
046	400 #d	610 #y	711 #f	811 #f	111 #d	511 #y	
100 #d	100 #f	611 #d	730 #d	830 #d	111 #y	530 #d	
110 #d	410 #d	611 #f	730 #f	830 #f	130 #d	530 #y	
111 #d	410 #f	611 #y	760 #d		130 #y	550 #y	
130 #d	411 #d	630 #d	762 #d		150 #y	551 #y	
240 #d	411 #f	630 #f	765 #d		151 #y	643	
243 #d	500	630 #y	767 #d		400 #d	644	
245 #f	505	650 #d	770 #d		400 #y	645	
247 #f	513 #b	650 #y	772 #d		410 #d	646	
260 #c	518	651 #y	773 #d		410 #y	667	
260 #g	523	655 #y	775 #d		411 #d		
263	533 #d	656 #y	776 #d		411 #y		
	534 #b	657 #y	777 #d		430 #d		
		69X	780 #d		430 #y		
			785 #d		450 #y		
			787 #d		451 #y		

Consistency in Choice and Form of Main Entry: A Comparison of Library of Congress and British Library Monograph Cataloging

Edgar A. Jones

Random samples of main entries on monograph catalog records created by the Library of Congress and the British Library were compared for the year 1989 in order to examine one aspect of the potential usefulness to the Library of Congress of British Library monograph cataloging, especially regarding the extent of agreement in cataloging practice between the two national bibliographic agencies. It was estimated that between 27.1% and 31.5% of the monograph entries in the printed 1989 annual cumulation of the British National Bibliography had been cataloged by both the British Library and the Library of Congress. It was further estimated that, for both choice and form of main entry, agreement was achieved between 60% and 70% of the time, and for choice of main entry alone, between 96% and 99% of the time.

The question of universal bibliographic control (UBC) is one that has occupied librarians for some time. As set forth by Dorothy Anderson in 1974, UBC

presupposes the creation of a network made up of component national parts, each of which covers a wide range of publishing and library activities, all integrated at the international level to form the total system.¹

At the national level, UBC requires legal deposit or similar legislation to assure receipt of each new publication at the national bibliographic agency (NBA), as well as a mechanism at the NBA to estab-

lish authoritative bibliographic records for these publications and to make those records available on a timely basis (both as individual records and collectively in a national bibliography). At the international level, UBC requires both recognition that the NBA in each country is the agency responsible for creating the authoritative record for that country's publications and the application of appropriate international standards to facilitate the exchange of such records.²

For historical and cultural reasons, the British Library has been more successful than the Library of Congress in meeting

Edgar A. Jones is a doctoral student at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Manuscript received July 17, 1991; revised November 23, 1991; accepted for publication December 2, 1991.

the national requirements of UBC, substantially conforming to the recommendations of the 1977 International Congress on National Bibliographies.³ At the same time, common adherence to the second edition of the *Anglo-American Cataloguing Rules* (AACR2)—a code based on international cataloging standards—has resulted to a large extent in both libraries meeting UBC's international requirements.

General compliance with the requirements of UBC ought to make British Library catalog records useful to the Library of Congress (and, by extension, to the American library community). Indeed, a 1987 survey of OCLC users suggested as much, with a majority of both public and medium-sized academic libraries finding British Library catalog records easy to use and considering their availability important.⁴

When catalog records, or elements of them, have been created by one organization according to criteria that make them virtually indistinguishable from records created by another organization (that is, the "error" rate is acceptably low), then the receiving organization can realize economies in cataloging costs arising from its ability to assign the handling of such records or elements to lower-level staff and apply the same quality-control procedures to these records as to those created internally. This is the principle underlying copy cataloging. Such records can then be used by the receiving organization as "building blocks" for its own cataloging, necessitating only the addition of such elements (call number, subject headings) as are needed for full integration into the local catalog.

It should be noted that in this sense the population of potentially useful British Library cataloging is circumscribed in at least two ways. First, the complexities of serials cataloging—compounded by differing conditions under which new serial catalog records are created between AACR2 and both the International Serials Data System (ISDS) and the 1988 edition of the *ISBD(S)* (*International Standard Bibliographic Description for Serials*)—make international cooperation in this area problematic.⁵ Second, consistency in the as-

signing of subject cataloging elements by these two agencies remains low, as shown in a recent study by Yasar Tonta.⁶

Main entries are crucial elements in descriptive cataloging, determining the effectiveness of the collocation function of the catalog, i.e., whether representations of a single work (whether indexed by main entry or analytical added entry) and related works (whether related bibliographically or by subject—a work "about" a work) will file in the same place in the catalog or be dispersed. While the model here is of a printed catalog, the principle applies with equal force to online catalogs, where it is a question of whether a single search argument will retrieve all representations of a work and works related to it, or whether an unknown number of arguments will be necessary.

BACKGROUND

Anglo-American cooperation in cataloging, broadly defined, began with the first Anglo-American code in 1908. This code and its eventual successor, the first edition of the *Anglo-American Cataloguing Rules* (AACR), while embodying a great many shared principles, still retained a sufficient level of disagreement among the parties that separate British and North American texts were deemed necessary.

The years between 1908 and 1967 were not marked by steady progress. While the 1908 code was applied in British libraries right up to the adoption of AACR in 1967, its North American version was superseded in the 1940s by a set of rules on choice and form of entry issued by the American Library Association and a set of rules on bibliographic description issued by the Library of Congress.^{7,8} These rules were superseded in turn by AACR. However, the Library of Congress, rather than implementing AACR *in toto*, chose instead to "superimpose" AACR practice on older practice, resulting, for example, in the retention of older forms of name headings until the adoption of AACR2 in 1981. Consequently, the implementation of AACR2 was much more convulsive in the United States than otherwise would have been the case.

The years immediately preceding the adoption of AACR2 saw dramatic changes in the world of library and information science. The most significant of these was the introduction of the MARC communications format by the Library of Congress, the subsequent development by other NBAs of analogous national MARC formats, and development of the UNIMARC format for international exchange. In addition, the spread of computer links among libraries and information centers dramatically changed the economics of cooperative cataloging. Bibliographic utilities, in exchange for access to large files of catalog records and to assure the integrity of a shared database, demanded rigid adherence to cataloging rules as well as to related national rule interpretations.

As more and more MARC records were generated, albeit following different national practices, demand increased that these disparate records be brought together as resources for all. At the level of the NBA, as earlier at the level of the local library, the implications were clear:

In an era when the exchange and the cooperative assembly of bibliographic information between national libraries and other agencies is assuming greater and greater importance as part of their role, the cost of mechanization means that the exercise of options and the divergences from the standards are luxuries that few people can now afford.⁹

In contrast to the two Anglo-American codes that had preceded it, AACR2 was published in a single text. American and British practice had moved much closer, though some disagreement remained in the form of optional rules and portions of rules. After implementation in 1981, the NBAs struggled to reduce their divergent application of the code, especially through the mechanism of ABACUS (Association of the Bibliographic Agencies of Britain, Australia, Canada, and the United States).¹⁰ The 1988 revision of AACR2 represented a significant move toward harmonizing practice.

Currently, cataloging cooperation between the NBAs consists of the conversion and distribution of one another's MARC records within their respective national

markets and, more important, adherence to a common cataloging code—AACR2—without which the former activity would be of limited use. Within cataloging databases in each country, foreign MARC records generally reside side-by-side with their home-grown analogs. For systems such as RLIN that routinely segregate records on the basis of inputting or tape-loading agency, such duplication simply reflects the agreed-upon practice. But even the OCLC Online Computer Library Center, Inc., a bibliographic utility that as a matter of policy generally eschews duplication, currently makes an exception for certain classes of MARC records created by national bibliographic agencies and allows multiple records to represent the same bibliographic item.

Although LC provides tape conversion services, it does not itself use foreign MARC records, and little effort has been made to control such records. Access points on foreign MARC records are run against the LC Name Authority File (LCNAF), but headings are converted to the LC form only when the foreign NBA form is identical to a reference on the relevant LC name authority record. For example, the British Library heading **Ordnance Survey** corresponds to the LC heading **Great Britain. Ordnance Survey**, but the nearest reference on the LC name authority record is from **Ordnance Survey (Great Britain)**. Such records are distributed with the original headings intact. Although LC's Shared Cataloging Division formerly accepted bibliographic descriptions from foreign national bibliographies, this practice was discontinued following the implementation of AACR2.¹¹

For now, the presence of British Library cataloging in OCLC and RLIN is the visible form and extent of Anglo-American bibliographic cooperation in the United States. Such duplicative cataloging appears contrary to common sense. Beyond this, the acceptance of its products into a bibliographic database is in violation of a principle underlying the development of the USMARC holdings format: that a single "universal" bibliographic record exist for a given bibliographic item in a given database.¹² The presence of multiple

bibliographic records poses problems for linked holdings records, which in turn inhibits interlibrary lending.

More active cooperation would appear inevitable. The Open Systems Interconnection (OSI) reference model promulgated as an international standard by the International Organization for Standardization (ISO) provides a mechanism for linking divergent computer systems for purposes of intersystem communication and record transfer. Widespread acceptance of the OSI reference model as the preferred method for achieving such intersystem linking in the area of library and information science has already resulted in the development and implementation of several national-level applications. Future cooperative cataloging programs between NBAs will likely use the OSI reference model as the environment within which communication and record transfer will take place.

In a North American context, the most significant functioning applications have been the Linked Systems Project (LSP) application for name authority records in the United States and the National Library of Canada's interlibrary loan protocol. An LSP application for bibliographic records is in the development stage and could be introduced within the framework of the National Coordinated Cataloging Program (NCCP), whose participants currently input directly to MUMS, the in-house cataloging database of the Library of Congress.¹³

The usefulness of OSI in the international exchange of bibliographic records will be problematic unless a mechanism is developed to support absolute convertibility of national MARC records. Due to differing repertoires of data elements, records cannot be converted back and forth between national formats without significant degradation of content-designation and, in some cases, content.¹⁴ For example, the UNIMARC format defines specific fields for bibliographic notes relating to each of the eight areas of the ISBD,¹⁵ while the UKMARC format is constructed to carry multilevel descriptions (see below).¹⁶ However, for one-way conversion into USMARC, as contemplated in

this study, this problem should not be insuperable.

LITERATURE REVIEW

A review of the literature since 1980 found little published research quantifying the difference in descriptive cataloging practice among NBAs within the AACR2 community. However, such research relating to AACR1 was performed by C. Donald Cook at Columbia University in 1977.¹⁷ Cook examined the descriptive cataloging of the Library of Congress, the British Library, and the National Library of Canada for a population of personal name headings used as main entries on three-member groups of bibliographic records. These records represented works by Canadian authors or works relating to Canadian subjects published in Great Britain and cataloged for *Canadiana* from 1968 to 1972 and any analogs from the *British National Bibliography* and the *National Union Catalog*. While one must obviously be cautious when extrapolating from such an unrepresentative subset of British and North American NBA cataloging, Cook's research did produce interesting results. His hypothesis was that

use of the *Anglo-American Cataloging Rules* by the Council of the British National Bibliography (now a unit of the British Library), the National Library of Canada, and the Library of Congress has not resulted in the standardization of choice and form of heading for the entry of works in the *British National Bibliography*, *Canadiana*, and the *National Union Catalog*.¹⁸

By virtue of the characteristics of its chosen population, Cook's study eliminated from consideration both government and serial publications and, by virtue of other necessary exclusions, reduced the remaining population of British imprints "by Canadians or . . . Canadian in content"¹⁹ to some 515 titles with 478 discrete headings used as entries. Cook then examined these items, categorized them as identical (16.1%), different in choice of entry (3.3%), and different in form of entry (83.8%) and found his hypothesis confirmed. He analyzed the nature of the

differences where they occurred and found that of the 83.8% with different forms of entry, 72.6% resulted from "following the code." He concluded that a mutually maintained authority file "may be the only means by which standardization might be achieved, regardless of the provisions which might be made a formal part of a code."²⁰

The elimination of "options" within AACR2 remains a focus of ABACUS as well as of the Joint Steering Committee for Revision of AACR, but no published studies explore the magnitude, in terms of incompatible bibliographic records, of the current problem. Cook analyzed corporate name headings used on AACR2 MARC records produced by the British Library, National Library of Australia, National Library of Canada, and Library of Congress in 1981, the first year of AACR2. This study, however, was limited to names entered directly and without subdivisions; thus again, while he produced results for a clearly defined population, the conclusions could not be extrapolated for the larger population of monographs, or even just for the larger population of corporate name headings.²¹

The National Library of Australia (NLA) also conducted research of this sort, producing a four-page study for the 1983 ABACUS meeting. The NLA study showed variations in 16.4% of cases compared with Cook's 5%. The unit of analysis in the NLA study, however, was the heading, while Cook's was the bibliographic record; the results are therefore not comparable.²²

It appears that some related work has been done internally at the Library of Congress—"an investigation . . . into the Library's potential use of MARC cataloging data from other national libraries"—but the study seems to have been limited to the question of whether or not to provide internal access to those records as "a Foreign MARC Resource File" at LC.²³

Finally, in a report on the work of the IFLA Section on Bibliography for 1987-88, "comparative research directed towards the compatibility of national bibliographies" is mentioned as being within the section's "terms of reference," though no current research is mentioned.²⁴

THE STUDY

The purpose of this research was to provide information that could assist in planning for mechanisms of inter-NBA cooperation and to contribute to the achievement of the larger goal of universal bibliographic control. Specifically, catalog records for monographs appearing in the 1989 annual cumulation of the *British National Bibliography* (BNB) were examined to determine the extent of agreement on choice and form of main entries with the corresponding cataloging from the Library of Congress.

To determine the potential usefulness of such records as building blocks for Library of Congress cataloging, two questions must be answered: (1) How many monographs are cataloged by both LC and the BL? and (2) What is the degree of consistency in choice and form of main entry for monographs cataloged by both institutions?

Differences in bibliographic description were not examined because the British Library and the Library of Congress have already determined that differences between them in this area are minor and "not significant in terms of record compatibility."²⁵

The sample examined—monographs for which both the British Library and the Library of Congress currently provide original descriptive cataloging—represented the population of most likely interest to any future US-UK cooperative cataloging venture.

It was expected that agreement in choice and form of entry, while still being far from total, would have increased in frequency since Cook's 1977 examination of AACRI practice.²⁶ This seemed reasonable, given the replacement of the separate British and North American texts of AACRI by a single AACR2 text and the abandonment in 1981 of the Library of Congress policy of "superimposition" and in 1982 of its policy of "compatibility."²⁷

METHODOLOGY

The population of interest were traditional main entries—the data recorded in

USMARC fields 1XX and 240, and subfields #a, #n, and #p of field 245—on a sample of 802 monographic bibliographic records, of which half represented British Library cataloging and half the corresponding Library of Congress cataloging.

For purposes of this study, "Library of Congress cataloging" refers to cataloging created either by the Library of Congress or by an institution participating with LC in a cooperative cataloging program. The rule options and interpretations applied and the authority and bibliographic files consulted are those of the Library of Congress. In the sample extracted, items in the medical sciences were for the most part cataloged by the National Library of Medicine; additionally four items each were cataloged by Harvard University and the University of Michigan.

LC online in-process records on OCLC (identified by the text "IN PROCESS (ONLINE)" in USMARC field 050) were included on the assumption that the descriptive cataloging had been completed to the same extent as for records created in the customary way.

To qualify for inclusion in the sample, it was necessary that the British Library records appear in the printed 1989 annual cumulation of the *BNB* and that the Library of Congress records have control numbers implying creation between 1987 and 1990. The first two numeric digits of an LC control number correspond to the last two digits of the year in which it is assumed work began on the record. This is necessarily an arbitrary assumption. On the one hand, control numbers are often assigned prior to the cataloging of an item, so that an item represented by a record with an "86-" prefix might actually have been cataloged in 1988. On the other hand, bibliographic descriptions, especially headings, are often revised subsequent to cataloging, so that a record with an "88-" prefix might contain elements that were revised in 1990. The 1989 *BNB* volume selected was the latest complete year available at the time of the study (fall 1990).

The 802 records were selected as follows. A photocopy was made of the printed 1989 annual cumulation of the *British National Bibliography* and the entries

numbered consecutively by hand. A random-number table was then used to select sample records.

It was determined that a sample size of 400 records (800 records paired) would be sufficient to make statements about the population with 95% confidence and intervals of 5% or less. One record that was rejected initially was subsequently found to be within scope, bringing the number of corresponding records to 401 (802 paired).

After eliminating from an initial sample of 1,699 entries those items cataloged as serials by the British Library (91 entries), the remaining entries were searched against the OCLC database to locate analogous LC monographic records. Of the 1,608 records searched, LC monographic cataloging was found for 471 entries, 70 of which were deemed out of scope according to the control number criterion described above and for other reasons. For example, in one case it was possible that a difference in the form of a personal name heading resulted from the "enhancement" of the LC record by an OCLC member institution.

To economize on space in printing the *British National Bibliography*, *BNB* entries do not always include all the data from the corresponding UKMARC records. Specifically, uniform titles are included only rarely, and until recently, dates and fuller forms of name were included in personal name headings only when necessary to distinguish otherwise identical names (an application of AACR2 rule 20.3). It was therefore necessary, for purposes of this study, to replace the printed *BNB* entries with USMARC records representing the corresponding machine-readable British Library cataloging. These records were also retrieved from the OCLC database. It was from these OCLC records that the data elements listed above were extracted.

Record pairs were characterized as "different" if they differed as to level of analysis (multilevel descriptions are used occasionally by the British Library), choice of entry, form of heading, or transcription of title proper.

To determine whether, for a given pair of records, meaningful differences existed in either the form of heading (when

present) or the transcription of title proper, these elements were considered identical if any differences detected would have been eliminated under the normalization procedures employed in the Linked Systems Project. In general, these procedures ignore capitalization, diacritical marks, and most punctuation, a notable exception being the comma separating the surname from following forenames in personal name headings.²⁸ In comparing records from different MARC formats of origin, normalization helps minimize differences arising out of the use of divergent character sets.²⁹

The 138 record pairs that contained meaningful differences were analyzed and the differences categorized according to the element and subelement involved (e.g., fuller form of a personal name) and, where appropriate, the AACR2 rule or NBA option or rule interpretation that had been applied. National applications of options and rule interpretations vary over time, and it was impossible in all instances to determine when a difference occurred because of national practices in these areas. The sources for determining differences of these types was Cook's "AACR2" *Decisions and Rule Interpretations*, 3d ed.³⁰ and, for LC rule interpretations, etc., the successive issues of *Cataloging Service Bulletin*.

The LCNAF on OCLC was searched to verify LC forms of heading and to provide information relating to the status of the heading vis-à-vis AACR2 (whether it was coded "compatible" and whether reference sources were used in its construction), the item on which it was used in the sample (whether that item was cited as the first work cataloged using the heading), and British Library Cataloguing-in-Publication (CIP) data (whether it had been used in the construction of the heading).

RESULTS

Library of Congress cataloging was found for 471 monographs out of a total random sample of 1,608 monographs cataloged by the British Library. Based on this result, it was estimated that such cataloging would have been found for 29.3 (2.2%) of the

monograph entries in the printed 1989 annual cumulation of the *British National Bibliography*. At the time of writing, figures had not been published for the total number of entries in the 1989 printed *BNB*. However, volumes for recent years have contained between forty and fifty thousand entries.

The entry pairs in the sample were evaluated for differences after being subjected to the LSP normalization criteria. Of the 401 pairs, 263 (65.6%) were categorized as being identical and 138 (34.4%) as being different. In terms of choice of entry alone, 391 pairs (97.5%) were identical and 10 pairs (2.5%) different. It was estimated from the sample results that, when cataloging printed books represented in the 1989 annual cumulation of the *British National Bibliography*, catalogers at the British Library and the Library of Congress arrived at identical choice and form of main entry 65.5% of the time ($\pm 5\%$) and that, in terms of choice of entry alone, agreement was achieved 97.5% of the time ($\pm 1.5\%$) (see table 1).

DISCUSSION

CATALOGING OVERLAP

That LC cataloging would have been found for between 27.1% and 31.5% of the entries in the printed 1989 annual cumulation of the *BNB* is a measure of the potential usefulness of British Library descriptive cataloging for the Library of Congress. Given that it is unlikely the Library of Congress had already cataloged all of the potentially pertinent monographs at the time the sample was taken (fall 1990), either because they had not yet been ordered or had been assigned a low processing priority, this estimate can be viewed as conservative.

It should be noted that many of the items cataloged by the two agencies technically are being published in both countries, a fact reflected by the presence in their imprints of places of publication in both countries. Therefore, while more than a quarter of *BNB* entries represent items acquired by the Library of Congress, a large percentage of these are acquired as U.S. imprints. In an attempt to judge the

TABLE 1
 AGREEMENT ON CHOICE AND FORM OF MAIN ENTRY,
 BY *BNB* CHOICE OF MAIN ENTRY

BL Choice of Entry	N	LC Choice and Form of Entry			
		Same		Different	
		f	%	f	%
Personal Name	281	154	55.0	127	45.0
Corporate Name	10	7	70.0	3	30.0
Uniform Title	1	1	100.0	0	0.0
Title Proper	109	101	92.7	8	7.3
Total	401	263	65.6	138	34.4

magnitude of this phenomenon, those items in the sample cataloged by the British Library at AACR2 level 2 (and therefore including the first place of publication) were analyzed. Forty-five and six-tenths percent (125 out of 274) listed a first place of publication in the United States. If a similar pattern were to hold for items cataloged at AACR2 level 1 and for items where the first place of publication is in Britain and a subsequent place in the U.S., then a large proportion of this overlap would represent domestic production.

On the other hand, the small overlap can be attributed to differences in selection criteria for inclusion in the *BNB* and the collections of the Library of Congress respectively. The British Library is a national library in the accepted sense, and as such it is responsible for providing, in the form of the *British National Bibliography*, a permanent record of the national imprint. To satisfy this obligation, the BL routinely catalogs whole classes of materials—paperback editions of items previously published in hardback, “mass market” paperbacks, school textbooks, children’s books, cookbooks, automobile maintenance manuals, and pamphlets—that it has no intention of adding to its own collections.³¹ The Library of Congress, on the other hand, is first and foremost a research library and acquires especially its foreign materials to satisfy the research needs of its primary clientele. Additionally, many items published in the United Kingdom are published simultaneously, earlier, or later by a different American publisher,

and it is this American edition that will most likely be acquired by LC.

EXTENT OF DIFFERENCE

The finding that nearly two-thirds of the record pairs were identical in choice and form of entry is a marked improvement over Cook’s *AACR* findings. There appear to be a number of factors that might have contributed to the observed improvement. First, the two editions of *AACR* were superseded by a single edition of *AACR2*. Second, with its implementation of *AACR2* in 1981, the Library of Congress abandoned its *AACR* policy of “superimposition.” Both of these events made it more likely that entries would be identical simply as a result of following the cataloging code.

Third, rules for choice of entry under *AACR2* result in more items being entered under title proper than was true under *AACR1*, and it was in this category that the highest degree of consistency (92.6%) was achieved. If works entered under title proper are excluded from the results, then the degree of consistency in the sample pairs drops from 65.6% to 55.6%.

Finally, the results also might have been influenced by a fourth factor: the presence of British Library CIP data in a large proportion of British publishing output. Through British Library CIP data, LC catalogers would have access on an item basis to British Library choice and form of entry (though this would not be true for items carrying both British Library and

Library of Congress CIP data). While this would not affect headings already established at LC, it would tend to encourage uniformity in such matters as the addition of fuller forms of name, and dates of birth and death in new personal name headings. However, examination of LC name authority records showed that, of 90 headings based on items represented in the sample, British Library CIP data were used to enhance the heading in only 1 case (by supplying a date of birth). In 71 cases, the BL and LC headings were identical; in 19 they were different. A breakdown of the 71 identical cases by the characteristics of the heading shows that, in more than 70% of the cases, the heading consisted of the name alone, unaugmented by other data (for all personal name headings in the sample, headings consisting of the name alone accounted for 72.8% of the identical cases)(see table 2). In the case of the differences, it was not possible to determine the extent to which the British Library headings were based on the items in the sample, though in two cases the form of the name differed from the form appearing in the chief source.

From this analysis it appeared that CIP data were not serving in any significant way as a surrogate for Cook's mutually maintained authority file. This may in part be accounted for by the large proportion of the overlap representing items published simultaneously in both countries. Such items would be candidates for both LC and BL CIP data. Consequently, the Library of Congress cataloger might receive the item as a publisher's galley proof without the British CIP data being present.

It should also be borne in mind, however, that the Library of Congress does not grant any special status to British CIP. Except for Canadian names, where a limited framework for cooperation is in place,³² LC's basic frame of reference remains its own name authority file.

DIFFERENCES IN CHOICE OF ENTRY

Ten entry pairs differed as to choice of entry. In the analysis that follows, they are grouped according to the type of access point serving as the *BNB* main entry.

TABLE 2
CHARACTERISTICS OF IDENTICAL
HEADINGS WHEN LC FORM IS BASED
ON A SAMPLE ITEM

Characteristics of Heading	N	%
Personal Name	50	70.4
Personal Name, Date	17	23.9
Personal Name (fuller form)	1	1.4
Personal Name (fuller form), Date	3	4.2
Total	71	99.9

Percentages do not total 100 due to rounding.

Of three items entered under personal name headings by the British Library, one was entered under the heading for a different person by the Library of Congress. In this case, the two agencies appear to have disagreed over the responsibility of a translator for the intellectual content of a work. The British Library entered the work under the heading for the individual whom the Library of Congress deemed simply to have translated the work. In another case, the editor was not identified prominently as such, and the British Library cataloger consequently mistook him for the author. In the third case, authorship changed between editions of a work, and the British Library cataloger was able to continue entering the work under the heading for the author of the earlier edition through the device of relegating the statement of responsibility to the edition area of the description.

In one case, the British Library entered under a conference name a publication that LC entered under title proper (the conference was named in the chief source).

The small number of bibliographic records entered under corporate name headings in the *BNB* sample (10 records out of a total of 401) can be attributed primarily to the action of AACR2 rule 21.1B2, which severely restricts the circumstances under which records may be so entered. In addition, several categories of material that would routinely be candidates for such entry (e.g., government publications; maps; annual reports, etc., of business firms) are explicitly or implicitly excluded from or

limited in the *British National Bibliography*. This leaves the field to conference publications, catalogs, inventories, etc. In the case of the 10 records in the sample, all represented conference proceedings.

In 6 cases the British Library entered under title proper items that the Library of Congress entered under other elements. Three of these were conference proceedings, one was a work produced under editorial direction, one was a work produced by a compiler, and one was a revised edition of a work for which the British Library cataloger apparently no longer considered the original author responsible. In 4 cases (all except the work produced under editorial direction and the work produced by the compiler), the Library of Congress entry resulted from applying the relevant AACR2 rule.

DIFFERENCES IN FORM OF PERSONAL NAME HEADING

Of 278 bibliographic records entered under a personal name heading where both the British Library and the Library of Congress agreed on choice of entry, 180 were entered under the identical heading by both agencies and 98 were entered under different forms of heading for the same person. Headings differed as to the fullness of the form of name serving as the basis of the heading (25 headings), the element of the name serving as the entry element (1 heading), and elements added to the name either to make the heading unique or for other reasons (75 headings).

When headings differed in the fullness of the form of name serving as the basis of the heading, the difference appeared in most cases to arise out of the differing cataloging contexts rather than out of differing applications of rule 22.3A1. That is, the "most commonly found" form of name was either different for the two libraries or had not become sufficiently preponderant in one library or the other to warrant revising the existing heading. However, in 8 out of the 25 cases, the difference seemed to be attributable, at least in part, to the policy of the Library of Congress to allow certain headings established prior to the implementation of

AACR2 in 1981 to be declared "compatible" with the new code even though in technical violation of it.³³ This was particularly true for many headings based on the full legal names of persons rather than on the forms under which they wrote (e.g., **Black, Clinton Vane de Brosse** rather than **Black, Clinton V.**). Additionally, two differences were attributable to the application by the Library of Congress of alternative rule 22.3C2, which allows the headings for persons whose names are written in a nonroman script and entered under surname to be based on the form of name appearing in English-language reference sources rather than on the form appearing most often in their works (**Dostoyevsky, Fyodor** rather than **Dostoevskii, F. M.**).

One heading differed in the element of the name serving as the entry element, with the Library of Congress deciding that the name (John Maynard Smith) consisted of a forename and a compound surname, while the British Library felt that it consisted of two forenames and a single surname.

Seventy-five headings differed in the elements added to the name either to make the heading unique or for other reasons. Sixty-one headings differed in the presence or absence of a date (in 35 cases this was the only difference between the headings); in 46 of these cases the British Library heading included a date, while in 15 cases, the Library of Congress heading did so. Two headings differed in the form of the date (1908-1986 vs. 1908- and b.1869 vs. 1869-1933). Thirty-three headings differed in the presence or absence of a parenthetical fuller form of name (in 12 cases this was the only difference between the headings); in 25 of these cases the British Library heading included a fuller form, while in 5 cases the Library of Congress heading did so. Four headings differed in the form of the fuller form (**John Raymond** vs. **John R.**, **Robert Harry** vs. **Robert H.**, **Timothy John Caldecott** vs. **Tim J. C.**, and **Thomas Patrick Joseph** vs. **Thomas P. J.**). One heading differed in the presence or absence of a distinguishing term. In this case, the Library of Congress added **MRCGP (Member of the Royal College of General Practitioners)** to the heading.

It should be noted that the more complex the *BNB* form of personal name heading, the less likely it was to be identical to the corresponding LC form, with unaugmented headings being identical 84% of the time and headings augmented by both fuller forms of name and date (the "worst case") identical only 15.6% of the time. Table 3 shows the percentage of headings found to be identical and different for each class of *BNB* heading.

DIFFERENCES IN FORM OF CORPORATE NAME HEADING

Of 9 bibliographic records entered under a corporate name heading where both the British Library and the Library of Congress agreed on choice of entry, 7 were entered under the identical heading by both agencies and 2 were entered under different forms of heading for the same body.

Seven headings were identical. It should be noted that, because all included local places as qualifying data, all seven headings would have differed prior to the 1988 revision of *AACR2*, when British and American practices with regard to rule 23.4B1 were reconciled. (American practice in such cases was always to include the name of a larger jurisdiction [e.g., **London, England**], while British practice was not to do so when it was not deemed necessary for identification [e.g., **London**]. The 1988 revision resolved the difference in favor of the American practice.)

Of two records entered under different forms of heading for the same corporate body, one heading was different due to a

transcription error by the Library of Congress cataloger rather than to a differing application of the rules, etc. The other differed due to the choice of a brief (acronym) form of name by the British Library (**MAFELAP 1987 (Conference : Brunel University)**) and a long form by the Library of Congress (**Conference on the Mathematics of Finite Elements and Applications (6th : 1987 : Brunel University)**), presumably through differing applications of rule 24.2D. Since the brief form has predominated in the body's publications in recent years, the LC heading would presumably be revised at some point.

DIFFERENCES IN FORM OF UNIFORM TITLE (INCLUDING UNIFORM TITLE MAIN ENTRIES)

In the printed *BNB*, the British Library makes less extensive use of uniform titles than the Library of Congress, but this is less the case with the corresponding UK-MARC records. While the *BNB* does not routinely display uniform titles for translations, the corresponding UKMARC records almost always include them. In all 9 such cases the British Library and Library of Congress uniform titles were identical. However, in certain cases of conflict between such a uniform title and one of another type, the Library of Congress will collocate by the uniform title in the original language, while the British Library will collocate by the uniform title in the language of the translation. There was one example of this in the sample, where a collection of short stories entered under

TABLE 3

CHARACTERISTICS OF PERSONAL NAME HEADINGS BY *BNB* FORM OF HEADING

<i>BNB</i> Form of Heading	N	LC Form of Heading			
		Same		Different	
		f	%	f	%
Personal Name	156	131	84.0	25	16.0
Personal Name, Date	72	40	55.6	32	44.4
Personal Name (fuller form)	18	4	22.2	14	77.8
Personal Name (fuller form), Date	32	5	16.6	27	84.4
Total	278	180	64.7	98	35.3

the heading for I. S. Turgenev used the uniform title [**Love and death**] on the British Library record and [**Short stories. English. Selections**] on the Library of Congress record.

Three records differed because the Library of Congress cataloger added the uniform title [**Selections**] subarranged by date of publication (LC RI 25.9A), while the British Library cataloger used no uniform title. One record differed in the presence of an initial article in the BL uniform title. One record differed because the Library of Congress cataloger applied rule 25.7A (using the uniform title for the first work for an item containing two works) while the British Library cataloger did not. One record differed because the Library of Congress cataloger applied rule 25.35E (making a uniform title for a libretto written by a composer) while the British Library cataloger did not.

DIFFERENCES IN TITLE PROPER

Aside from typographical errors, titles proper differed primarily in their extent, with the British Library tending to treat as other title information what the Library of Congress tended to include in the title proper. For example:

<i>BNB title proper</i>	<i>LC title proper</i>
Hemingway	Hemingway, the Paris years
The atheist	The atheist and other short stories
Ira Hayes	Ira Hayes, Pima marine
Dublin	Dublin, one thousand years

Seven pairs fell into this category. Additionally, one pair differed in the representation of the Greek character alpha, which the Library of Congress cataloger

transcribed as [alpha] while the British Library cataloger transcribed it as the character. This apparently reflects the application of LCRI 1.0E1, which directs that all Greek characters be recorded in this manner, although some, including alpha, are represented in the ALA character set.

DIFFERENCES IN LEVEL OF ANALYSIS

The British Library applies AACR2 rule 13.6 (multilevel description) to many items published in separate parts, while the Library of Congress treats such works either as separate monographs or as unanalyzed parts of multivolume monographs. In converting British Library records to the USMARC format, the Library of Congress converts the item as a representation of the first-level data (pertaining to the whole), with second-level data (pertaining to the part) relegated to a partial contents note. As a result, when records for different parts of an item are converted, they appear to be duplicates of one another, differing only in the content of the partial contents note and those other elements (e.g., uniform title) that may relate to it. Five pairs fell into this category. The BNB record represented in figure 1 exemplifies the complexities.

TYPOGRAPHICAL, ETC., ERRORS

Twelve records contained typographical and other extraregular errors. Eight entries contained mistranscriptions of the title proper (7 with words missing, 1 with differing spellings of "behavior"). One British Library record entered under a personal author contained a PRECIS string in place of the title proper. Two records—one from each NBA—contained mistranscriptions

Bataille, Georges, 1897-1962. [La part maudite. 1. English] The accursed share : an essay on general economy / Georges Bataille. - New York : Zone ; London : Distributed by MIT. Translation of: La part maudite, vol. 1 Vol. 1: Consumption / [translated by Robert Hurley]. - 1988. - 197 p. ; 24cm
ISBN 0-942299-10-8 (cased) : £18.95
ISBN 0-942299-11-6 (pbk) : no price B89-16612

of name headings. Finally, one error involved a BL misspelling of a word in a uniform title.

CONCLUSIONS AND RECOMMENDATIONS

CATALOGING OVERLAP

As noted earlier, the cataloging of the British Library might serve as a building block for the corresponding cataloging of the Library of Congress (and vice versa) given a large enough population of items cataloged by both libraries and a sufficiently high level of consistency.

A sizeable overlap was found between the cataloging of the British Library appearing in the *BNB* and that of the Library of Congress. Is the overlap large enough to justify the use of British Library descriptive cataloging data by the Library of Congress?

The overlap represented both purely British imprints and a sizeable population of dual British and U.S. imprints. While there can be no question that British Library descriptive cataloging data would be useful for the former class of publications, there are questions regarding the latter class. Because they represent items that can legitimately be considered to fall within the jurisdiction of both NBAs in the context of UBC, some understanding would have to be reached regarding which agency had priority in supplying descriptive CIP data in each case. This might be done most easily by assigning to each agency the production of a given set of publishers. Subject analysis and other categories of NBA-specific CIP data might still be supplied by the other agency. Because both NBAs give a high priority to processing CIP items, deferral to one agency or the other in this fashion should not result in any significant reduction in timeliness. Should the two agencies enter into such an arrangement, the savings in descriptive cataloging activity to the Library of Congress might constitute the whole of the class of purely British imprints and from one-third to one-half of the class of dual imprints, perhaps as many as 10,000 records.

A larger question, however, concerns

the British Library's 1987 decision to catalog 50% of the items appearing in the *British National Bibliography* at AACR2 level 1 (based on an analysis of user needs and the uses to which the records were actually put). The categories of material that were to make up this 50% were modern English fiction; children's books; material with 32 pages or fewer; and works on science, technology, and religion (Dewey 500-599 and 200-299).³⁴ Records in this class constituted 31.7% of the cases in the 1989 *BNB* sample. In contrast, the Library of Congress catalogs all currently received monographs at AACR2 level 2.

Hope Clement has noted the problems that reduction in cataloging standards pose to the achievement of UBC.³⁵ As NBAs come under economic pressure to increase output through reducing standards, the danger of divergent minimal standards increases.

In the present instance, there are three immediately apparent ways to resolve this divergence. First, the British Library might rescind its level 1 policy (for economic reasons alone this course seems unlikely). Second, both libraries could continue to follow their respective practices, with the Library of Congress simply augmenting the BL level 1 records with additional level 2 data (though such an exceptional policy would almost certainly result in a degradation of potential cataloging efficiency). Third, the Library of Congress and the British Library might negotiate between themselves a common categorization of records to receive level 1 cataloging, implicitly recognizing that, even at the national "authoritative" level, such categories may be justifiable. This would also necessitate agreement on what elements constituted level 1 cataloging, because for its "minimal-level" cataloging the Library of Congress currently provides somewhat more data than is strictly required by AACR2 rule 0.1D1.

EXTENT OF DIFFERENCE

A high incidence of identical choice and form of entry between the British Library and Library of Congress was found.

Although an exhaustive examination of the application of AACR2 by these two agencies is not a part of this study, the sample results do indicate that little would be gained from further refinement of the rules themselves. On the other hand, in the area of NBA policy, benefits would certainly result both from further agreement on the application of AACR2 chapter 25 (uniform titles) and from abandonment by one agency or the other of its practice relating to rule 22.3C2.

The greatest gain, however, would occur through recourse to a common name authority file (as Cook had found for AACR1). Use of such a file would presuppose the revision by the Library of Congress of its remaining "AACR2-compatible" headings as well as development of a one-time method for resolving differences between headings already established by the two agencies (e.g., favoring the form used by the NBA of an author's country of residence, etc., except when this form would conflict with another heading in the joint authority file).

The OSI reference model would presumably be the mechanism through which such an authority file would be developed. It should be noted, however, that that model has so far been applied only in a context of records that are virtually identical in structure (e.g., USMARC). It has yet to be seen how it would function in a context of USMARC and UKMARC records, where, of necessity, conversion from one format to the other would have to occur whenever any record transfer took place. This should not, however, be an insuperable obstacle.

Cooperation between the British Library and the Library of Congress—the world's largest creators of AACR2 cataloging records—in developing a joint database would be a logical first step on the road to a functioning system of UBC based on national NBA responsibility for the national imprint. This study has shown that, in terms of the size of the cataloging overlap between these two agencies and the magnitude and nature of the differences in their descriptive cataloging practice, there exists a strong basis for developing such cooperation.

REFERENCES AND NOTES

1. Dorothy Anderson, *Universal Bibliographic Control: A Long Term Policy, A Plan for Action* (Pullach/München: Verlag Dokumentation, 1974), p.11.
2. Ibid.
3. National Bibliographies Seminar (1987: Brighton, England), *Proceedings of the National Bibliographies Seminar: Brighton, 18 August 1987*, ed. Winston D. Roberts (London: IFLA Universal Bibliographic Control and International MARC Programme, 1988), p.73-75.
4. V. Louise Saylor, "UK MARC Records—Boon or Bane," *Technicalities* 8, no.10:7-9 (Oct. 1988).
5. ISBD(S): *International Standard Bibliographic Description for Serials*, Rev. ed. (London: IFLA Universal Bibliographic Control and International MARC Programme, 1988), p.2.
6. Yasar Tonta, "A Study of Indexing: Consistency between Library of Congress and British Library Catalogers," *Library Resources & Technical Services* 35:177-85 (1991).
7. A.L.A. *Catalog Rules: Author and Title Entries*, prelim. American 2d ed. (Chicago: American Library Assn., 1941); *A.L.A. Cataloging Rules for Author and Title Entries*, 2d ed., ed. Clara Beetle (Chicago: American Library Assn., 1949).
8. *Rules for Descriptive Cataloging in the Library of Congress* (Washington, D.C.: Library of Congress, 1949).
9. Peter R. Lewis, "The Politics of Catalog Code Revision and Future Considerations," in *The Making of a Code: The Issues Underlying AACR2*, ed. Doris Hargrett Clack (Chicago: American Library Assn., 1980), p.7.
10. The first ABACUS meeting (before the group had a name) was reported in the *Library of Congress Information Bulletin* 36:239-52 (1977); the second and fourth through sixth were reported in the *British Library Bibliographic Services Division Newsletter* 11:5-10 (1978); 20:22-23 (1981); 22-23:6-7 (1981); 28:3-5 (1983); [the sixth was also reported in *National Library Technical News* 1983, no.1:1-4]; the seventh was reported in *Acquisitions, Bibliography, Cataloguing News* 13:4-8 (1984). More recent meetings have not been publicly reported.
11. "Shared Cataloging," *Cataloging Service Bulletin* 17:32 (1982).
12. *USMARC Format for Holdings Data: Including Guidelines for Content Designation*

- (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1989), p.vi.
13. *Open Systems Interconnection: The Communications Technology of the 1990's: Papers from the Pre-Conference held at London, August 12-14, 1987*, ed. Christine H. Smith (New York: Saur, 1988).
 14. Ellen Gredley and Alan Hopkinson, *Exchanging Bibliographic Data: MARC and other International Formats* (Chicago: American Library Assn., 1990), p.283-86.
 15. *UNIMARC: Universal MARC Format* (London: IFLA International Office for UBC, 1977), p.42-43.
 16. Gredley and Hopkinson, *Exchanging Bibliographic Data*, p.130-66.
 17. Charles Donald Cook, *The Effectiveness of the Anglo-American Cataloging Rules in Achieving Standardization of Choice and Form of Heading for Certain Library Materials Cataloged in Canada, Great Britain, and the United States from 1968 through 1972* (D.L.S. diss., Columbia University, 1977).
 18. *Ibid.*, p.8.
 19. *Ibid.*
 20. *Ibid.*, p.165.
 21. C. Donald Cook, "Headings for Corporate Names: International Standardization under AACR2," *Library Resources & Technical Services* 28:239-52 (1984).
 22. "ABACUS 1983," *Acquisitions, Bibliography, Cataloguing News* 13:5 (1984).
 23. Library of Congress, *Annual Report of the Librarian of Congress for the Fiscal Year Ending September 30, 1987* (Washington, D.C.: Library of Congress, 1988), p.62, 78.
 24. B. C. Bloomfield, "Report of the Work of the Section on Bibliography 1987-88," *International Cataloguing and Bibliographic Control* 17:54-55 (1988).
 25. "ABACUS 1983," *Acquisitions, Bibliography, Cataloguing News* 13:5 (1984).
 26. Charles Donald Cook, "The Effectiveness of the Anglo-American Cataloging Rules."
 27. "Compatible Headings," *Cataloging Service Bulletin* 17:31 (Summer 1982).
 28. "LSP Normalization," *Cataloging Service Bulletin* 32:58-59 (Spring 1986).
 29. Gridley and Hopkinson, *Exchanging Bibliographic Data*, p.140, 149.
 30. C. Donald Cook, "AACR2" *Decisions and Rule Interpretations*, 3d ed. (Ottawa: Canadian Library Assn., 1985).
 31. *The British Library Bibliographic Services Newsletter* 44:1-3 (1987).
 32. "Canadian Headings," *Cataloging Service Bulletin* 22:42-44 (Fall 1983).
 33. "Descriptive Cataloging: Library of Congress Rule Interpretations," *Cataloging Service Bulletin* 18:49-51 (Fall 1982).
 34. *The British Library Bibliographic Services Newsletter* 44:2 (1987).
 35. Hope E. A. Clement, "National Bibliographic Agencies Cataloguing Survey," *International Cataloguing and Bibliographic Control*, 19:9 (1990).

Variations in Personal Name Access Points in OCLC Bibliographic Records

Arlene G. Taylor

The kinds of variations appearing in name access points found in OCLC bibliographic records were explored by examining a sample of records taken from the OCLC Online Computer Library Center, Inc., Online Union Catalog. Name access points were searched in the Library of Congress Name Authority File (LCNAF), and all records for each name were examined in the OCLC bibliographic file to identify variant forms. For 64.4% of the sample records, an authority record was found for one or more names on the record. Of these, 23.4% (15.1% of the entire sample) had one or more names that differed in form from the LCNAF form. Another 5.8% of sample records had personal names that had no LCNAF record but differed from the majority of records using that name. An average of 3.4 forms of access points per name was found. Prolific authors figured prominently. Only 38 names accounted for 84.5% of all records examined for the 457 personal names. Nearly 44% of variants were identified as being a "near match" to the standard form. Nearly 29% were found to be only a single typographical error away from matching the standard. Over two-thirds of all variations occurred in dates. It might be possible to correct many variations with machine assistance. Single typographical errors and near matches could be identified by machine for human review.

During the years preceding the advent of the *Anglo-American Cataloguing Rules*, second edition (AACR2), the concept of authority control for names was seldom discussed, and the need for it was considered debatable, even though it had been strenuously practiced at the Library of Congress (LC) for many years. With

implementation of AACR2, libraries that had formerly assumed that LC would provide them with name access points in consistent forms realized that they would have to do their own authority work if they wanted consistency in local catalogs. (Inconsistency in name access points on LC records has always existed to some

ARLENE G. TAYLOR is Associate Professor, School of Library Service, Columbia University. The research reported in this paper was supported by a grant from the OCLC Online Computer Library Center, Inc., Library School Research Equipment Support Program. A version of this paper was submitted to the OCLC Office of Research as a project report. A much shorter version was presented to the LITA/ALCTS Authority Control in the Online Environment Interest Group at the ALA Annual Conference, June 24, 1990. A very short report of preliminary results was published in the *Annual Review of OCLC Research* (July 1987-June 1988), p. 53-54. Manuscript submitted November 19, 1991; accepted December 20, 1991.

extent, but before AACR2, many libraries either handled it without authority files or ignored it.)

By 1979 the idea was considered important enough in the United States to merit its own conference,¹ and in the past ten years research and writing in the library literature on authority control have increased dramatically.² Oddly, there has been little consideration of the concept of authority control of names in the literature of information retrieval, although there has been a great deal about retrieval of relevant documents by subject, including research on the efficacy of controlled vocabulary versus keyword approaches. Michael Buckland has observed:

It is important to emphasize that "subject" access is only one example of retrieval. Documents and data can have "contextual" attributes assigned to them for the purpose of retrieval, such as author, publisher, and date of creation. Indeed, in academic libraries, more use is made of author entries in the catalogs than of subject entries.³

Later in the same work Buckland states:

Writings about retrieval and especially about the evaluation of information retrieval systems have been dominated by just one of the apparently unlimited range of attributes: subject matter, i.e., what documents are about.

Retrieval using the attribute of what documents are about . . . has dominated so much that it has, perhaps, hindered clarity of thought about the foundations of information retrieval theory. . . . Our conceptual framework and definitions should be broad enough to include all attributes, not just one.⁴

Even though there has been little research in information retrieval on the authority control of names, citations have been used for a number of years to retrieve related documents. The assumption is that the cited document is related to the source document in which the citation appears.⁵ Citations are typically searched by author or title. The assumption is made that the form of name used in the citation will allow it to be found in an index or catalog. But citation practice is widely variant—sometimes using the name as it is found on the document, and sometimes abbreviating it

drastically. A series of articles in 1985 and 1986 addressed the problems of searching for names in uncontrolled online databases.⁶ Actual usage of names by authors can also vary. Elizabeth Fuller found that 17.6% of a sample of personal authors taken from a library catalog used more than one form of name in works found in the catalog (i.e., not including journal articles, chapters in books, etc.).⁷ Tamara Weintraub found a comparable figure of 18.5% in a later study.⁸ The larger the file the more likely such varying forms are to be separated from one another. Yet any searcher who looks for the name of a person in a catalog or index surely expects both high recall and high precision—that is, the searcher expects to find *all* documents that relate to the person and to find *only* documents related to that person. After all, if the citation sought is likely to be relevant to one's needs, might not other works by that person also be relevant? And wouldn't works by another person with the same undifferentiated name be confusing?

While most commercially produced indexes still do not provide authority control for names, vendors of online catalogs serving libraries began to provide various degrees of authority control almost all at once in 1984.⁹ The major networks had moved somewhat earlier to address the problem; Western Library Network (WLN) and UTLAS have provided some form of verification of names since the implementation of AACR2; OCLC verified names in the bibliographic file against the authority file through a batch process just prior to the implementation of AACR2 and again in the spring and summer of 1987. OCLC and the Research Libraries Information Network (RLIN) have provided online access to LC's Name Authority File (LCNAF) for a number of years, but there has been no linkage between bibliographic and authority files. As a result, typographical errors, changes of forms of names by LC, and names taken from chief sources of information that vary for the same person or corporate body have not been made to conform to a consistent usage in OCLC and RLIN. Upon completion of OCLC's 1987 authority file matching project some, but not all, of the inconsistencies had been corrected.

Even when the LCNAF, which currently includes records for names, uniform titles, and series, is linked to a bibliographic file, all inconsistencies cannot be resolved. Research has shown that at least 52.7% of LCNAF records contain no references.¹⁰ Yet many of those records represent names that appear in variant forms in a bibliographic file. This can happen because of error on the part of the inputter or because a particular library has found a variant form of name on the chief source of information of a work to which LC does not have access. Even when references are present there are variants not covered by the references.

RESEARCH OBJECTIVES

This study was conducted to determine the extent to which OCLC bibliographic records contain variants in personal and corporate name access points, to learn whether any variants are covered by references in the LCNAF, and to categorize the variants in order to determine what methods of programming could be used to assist in providing consistency.

The research questions addressed were:

1. What proportion of records in the OCLC database has name access points for which LCNAF records may be found? How often do the name access points and the authorized LCNAF forms agree exactly? How often do the name access points agree with references on the LCNAF records?
2. How often do name access points found on sample OCLC records conflict with access points for the same name on other records in the database? Are variants input before or after appearance of corresponding LCNAF records, and what types of libraries input them? In what ways do access points for a name vary from the standard form for that name?
3. Are there variants in access points that could be corrected by a computer program? Are there variants that could be found by a computer program and then be corrected after human review?

METHODOLOGY

A sample of bibliographic records was drawn from the OCLC bibliographic file. Drawing a sample of records, rather than a sample of names, means that there is a very large representation of prolific authors and prolific corporate sponsors, because these have a much greater chance of having one of their records drawn than do authors or corporate bodies with only one or two records in the file. This, however, is valid because libraries usually deal with records, not individual names, and need to know what proportion of records may contain one or more names that vary from authorized forms. Libraries also deal with records containing access points for many prolific authors and corporate bodies in any given period of time.¹¹ In addition, one way to initiate a database "cleanup" project is to start at some point with the next record added, resolve all conflicts found for names on that record, and then proceed to the next record. Another way to proceed is to clean up all prolific authors first, as these represent a very large proportion of the records in a database and thus will make the greatest impact upon increasing both recall and precision in database searches.

Sample size was determined using the formula $n = (z/e)^2(p)(1-p)$, where e is the error level of precision, z represents the curve value for the confidence interval, and p is the probability that something will occur (in this case, the probability that one name access point on a record will disagree with a standard form for the access point). A small pilot test indicated that 18% ($p=.18$) of access points would disagree with a standard form. For a confidence of 95% ($z=1.96$) and a precision of .025, a sample size of 929 records was required. The sample was drawn at OCLC through use of a random-number program that pulled the records using OCLC record numbers. Because of the large number of prolific authors, it became apparent that the research would take many years if all 929 records were used. Therefore, after consultation with research staff at OCLC, it was decided to use the first 450 records,

which yields a precision of .03 with a confidence level of 90% ($z=1.65$).

"Authority work" was completed for every nonsubject personal name access point (i.e., MARC 100 or 700 field) and for every nonsubject corporate name access point (i.e., MARC 110, 111, 710, or 711 field). Where possible, authority records were found in the online LCNAF, and then the bibliographic file was searched, looking for records that related to the same name. Research assistants helped with much of the searching process, but it was discovered that there were so many records for some names that looking at each of them manually would have added at least a year to the project. Therefore, 33 names were sent to Edward T. O'Neill at OCLC, who wrote computer programs to search for records for these names. The resultant files, consisting of the record number and heading line from each record the computer had identified as possibly being the name in question, were then examined. Additional manual searching was done for all names based on experience that showed what kinds of errors often occur in the first four letters of surnames and in first letters of first forenames. For each variant, whether found manually or by computer search, research assistants printed out the record for further examination and coding. Variant forms of each name found in both access points and statements of responsibility were noted, and catalogers' methodologies were used to determine which variants represented the same person or body.

For each name the following data were collected:

- Presence or absence of an LCNAF record, and when an LCNAF record was found, agreement or disagreement of the sample name access point with the authorized form
- Total number of records for the name in the database
- Number of different access point forms found for each name
- Number of records in which the form of the name access point varied from the standard ("Standard" was defined as the form authorized by LC when an LCNAF record was found, or the

form that appeared on the largest number of records when there was no LCNAF record. The latter part of this definition resulted from the desire to investigate how much variation could be handled by a computer. A computer program could not judge which of two or more variant forms would be the correct AACR2 form.)

- Date of input, rules coding, and number of references on the LCNAF record, if present

For each variant identified, data were collected in the following categories:

- Type of variation
- Presence and nature of subfield "w" (*w) contents
- Source of input of record (e.g., LC, UKM, OCLC member)
- Date of input of record
- Encoding level of record
- Whether variant was an exact or near match to an LCNAF reference, a near match to the authorized heading, or had a single typographical error and otherwise would have been an exact match to the heading ("Near match" was defined as one in which one form is wholly contained within the other and there is no conflicting information. This definition was the result of the desire to be able to program a computer to identify near matches.)

FINDINGS

Of the 450 sample records, 25 (5.6%) have no personal or corporate name access points. The remaining records yield 457 personal names on 348 records and 153 corporate names on 133 records. Personal names only are found on 292 records, 77 records have corporate names only, and a mixture of personal and corporate names is found on 56 records.

Table 1 shows that authority records were found for 275 of the personal names (60.2%) and for 126 of the corporate names (82.4%). Two-hundred-ninety bibliographic records (64.4%) have one or more names for which an authority record was found in the LCNAF. For 247 of these (54.9% of the sample) an authority record was found for every name on the record.

TABLE 1

VARIATIONS BETWEEN FORMS OF ACCESS POINTS FOR PERSONAL AND CORPORATE NAMES FROM SAMPLE RECORDS AND FORMS FROM LCNAF RECORDS

	Records	Records with No Names	Personal Names on Remaining Records	Corporate Names on Remaining Records
Number in Sample	450	25 (5.6%)	457 (on 348 records)	153 (on 133 records)
Authority Records Found	290 (64.4%)*		275 (60.2%)	126 (82.4%)
Number Names that Differ from Form Found on Authority Record	68 [†] (23.4% of records with one or more authority records; 15.1% of all records in sample)		50 (18.2% of personal names with authority records)	20 (15.9% of corporate names with authority records)

*This number represents the records for which an authority record was found for at least one name on the record. For 247 records (54.9%) an authority record was found for every name on the record.

[†]This number represents the number of records for which at least one name on the record does not match the authority record for that name. Two records have two names each that do not match, making a total of 70 names in the sample that do not match the respective authority records.

Of the 290 records, 68 have one or more names that differ in form from the form on the corresponding authority record. This represents 15.1% of all records in the sample or 23.4% of the bibliographic records for which one or more corresponding authority records were found.

Because the sample is a sample of records, not of names in the database, the data for names can be taken only as the literal percentages of the sample. Of the 275 personal names with authority records in the sample, 225 match the corresponding authority record exactly. This represents 81.8% of the personal names with authority records. Fifty of these names (18.2% of the personal names with authority records) do not match in form. Of the 126 corporate names with authority records, 106 (84.1% of the corporate names with authority records) match the corresponding authority record exactly. Twenty of these names (15.9% of the corporate names with authority records) do not match in form.

In the remainder of this paper the data concerning personal names are discussed. The data for corporate names will be addressed in a later paper. In the findings that follow, data concerning personal names that have LCNAF records are presented separately from those that apply to per-

sonal names without LCNAF records, and all further reference to "names" refers to personal names.

Table 2 presents basic data for the two groups of names, with and without authority records. The effects of sampling from a file with prolific authors can be seen clearly. Previous research has posited that two-thirds of authors appear only once in library catalogs. Actually, I believe that proportion would not hold true in the OCLC file because of the effect of the relatively new phenomenon of multiple manifestations—that is, many authors may write only one work, e.g., a dissertation, but it in turn is microfilmed in film and fiche versions, and when those are cataloged, there are three bibliographic records. Even taking this into account, an overall average of 147.7 records per name is definitely not expected. Looking at the difference between the two groups, however, it is clear that authority records have been made for the names that appear most frequently and thus would seem to need authority control the most. Despite this there are more different forms per name (4.5 versus 1.8) in the LCNAF group than in the "no LCNAF" group. As we shall see later, this reflects the many additional opportunities to make errors in inputting the name.

TABLE 2

BASIC FINDINGS ABOUT THE GROUP OF PERSONAL NAMES FOR WHICH AUTHORITY RECORDS WERE FOUND COMPARED WITH THE GROUP OF PERSONAL NAMES FOR WHICH AUTHORITY RECORDS WERE NOT FOUND

	LCNAF Record Found		No LCNAF Record Found	
Number Names ¹ from Sample Records	275	60.2% of names in sample	182	39.8% of names in sample
Number Names from Sample Records Different from Standard ¹	50	18.2% of names with LCAF records	26	14.3% of names without LCAF records
Total Number Records Examined for Form of Name	65,902	97.6% of all records examined (239.6 records per name)	1,616	2.4% of all records examined (8.9 records per name)
Number Different Forms Found	1,248	4.5 forms per name	319	1.8 forms per name
Number Records in Total Group with Form Different from Standard	4,205	6.4% of total number of records found for names in this group	281	17.4% of total number of records found for names in this group

*"Names" in this and following tables refers to personal names only.

¹"Standard" in this and following tables refers either to the form authorized by LC when an LCNAF record was found, or to the form that appeared on the largest number of records for the same name when there was no LCNAF record.

Because of the unexpected disparity in the LCNAF group between the percentage of names from sample records that differ from the authorized form (18.2%) and the percentage of names from all the records in this group that differ (6.4%), the data were examined for an explanation. The prolific authors seem to have affected this outcome. Table 3 shows the comparable figures when data for the 38 names with more than 200 records are separated from the rest. It can be seen that for the 237 names with fewer than 200 records, 17.4% of records for those names bear forms different from those in the LCNAF records. This figure is much closer to the 18.2% of names from the sample records that differ from the forms on their LCNAF records. A t-test of the two samples indicates that the difference is not statistically significant ($t=0.34$). On the other hand only 4.7% of the records for the 38 names with 200 or more records bear forms different from those in the LCNAF records. This difference is statistically significant ($t=10.55$). Why would there be propor-

tionately fewer variants for famous authors? Perhaps it is *because* they are famous and so their names are better known and are less frequently misspelled. There is also the fact that the searching for the majority of these was done by computer rather than manually. While some variants that the computer did not pick up were found through manual searching, there could have been more. Even though there are proportionately fewer variants for the highly prolific names, they still account for more than half of the variants. The 38 names (8.3% of the names studied) are responsible for 59.4% of all the variants found. It can be seen that cleanup projects involving these names would be quite productive in terms of elimination of variants from the system.

CHARACTERISTICS OF RECORDS FOR VARIANTS

The records for all variants that were found in the project were analyzed for certain characteristics. Table 4 identifies input

TABLE 3

BASIC FINDINGS ABOUT THE GROUP OF PERSONAL NAMES FOR WHICH AUTHORITY RECORDS WERE FOUND, SEPARATING AND COMPARING NAMES WITH 200 RECORDS OR MORE AND NAMES WITH 199 RECORDS OR LESS

	200 Records or More		199 Records or Fewer	
Number Names from Sample Records	38	13.8% of names for which LCNAF record was found	237	86.2% of names for which LCNAF record was found
Number Names in Sample Different from LCNAF Record	5	13.2% of names in this group	44	18.6% of names in this group
Total Number Records Examined for Form of Name	57,078	84.5% of all records examined (1,502.1 records per name)	8,824	13.1% of all records examined (37.2 records per name)
Number Different Forms Found	608	16.0 forms per name	640	2.7 forms per name
Number Records in Total Group with Form Different from LCNAF Record	2,666	4.7% of total number of records found for names in this group	1,539	17.4% of total number of records found for names in this group

sources of the records and the percentages of the LCNAF group that were entered after the LCNAF record was available. As expected, two-thirds to three-fourths were input by OCLC members. Some users of OCLC may be surprised to learn that only 8.9% of the variant records were input by the British Library (UKM). The reason for the perception that UKM records vary more often may relate to the fact that several years' worth of UKM records were loaded into the system at one time in the early 1980s. The number of records encoded "I" (input to highest standards) or "blank" (records of national libraries) were counted because it was thought that perhaps more errors would be found on records encoded "K" (input to a lower standard), "7" (minimal-level cataloging records), "8" (cataloging-in-publication records), etc. This was not the case.

Also counted were the records in the "LCNAF found" group that were entered after the last revision of the LCNAF record. It is not possible to tell what was done when an LCNAF record was revised—in some instances the form of authorized heading may have been changed. Because

of this, only records entered after the last revision could definitely be counted as inputting errors. Most revisions to LCNAF records, however, are for the purpose of adding references. Therefore, there are likely more of the variants that were input incorrectly after the current correct heading was available than are shown in table 4. Even without counting this possible group, it is disappointing to see as much as 15% entered incorrectly when the correct form was available. These should be more certainly avoided with OCLC's new PRISM service, which allows copying of a heading from the LCNAF to the bibliographic record by machine rather than manually. These are also avoided in a system that uses the authority record number in the heading field and then displays the heading from the authority record when the bibliographic record is displayed.

Further analysis of date of input is encouraging, however. Records with variants were grouped by date of input and then compared with the proportions of records input into the whole database for the same year groupings. If variants had been input at a steady rate, one would expect the pro-

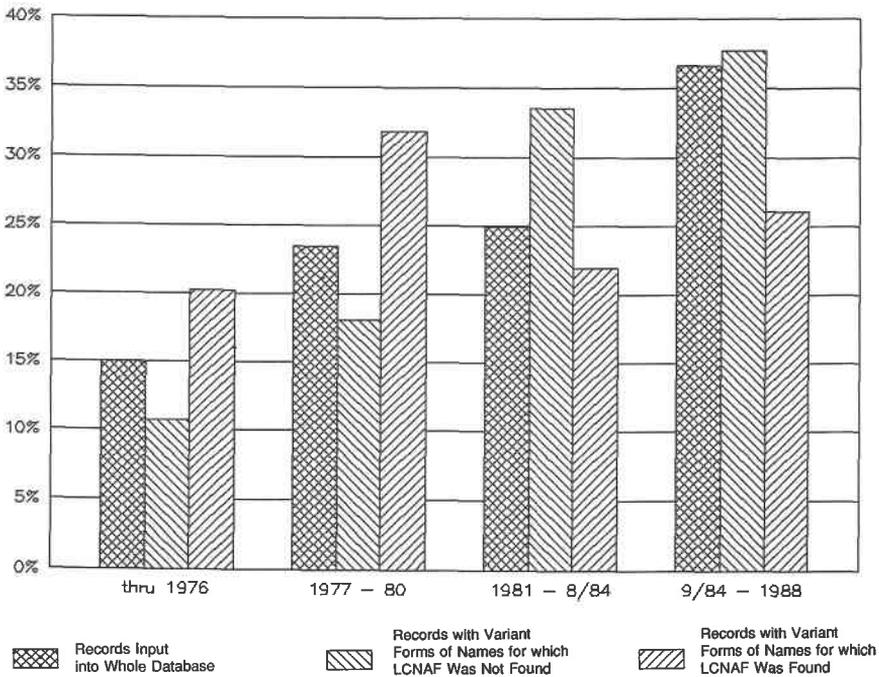


Figure 1. Comparison of Percentages of Records Input into the OCLC Database in Certain Groups of Years with Percentages of Records with Variants Input during Those Years.

portion of variants input in a given time period to be equal to the proportion of the whole database input during that time period. Figure 1 shows the results of the comparison. The break between 1980 and 1981 was chosen because of the implementation of AACR2 in January 1981. The

breakpoint of August-September 1984 was chosen because the current version of the LCNAF became available through OCLC in August 1984, although some authority records had been available online since 1980. No one date can give an accurate picture of the availability of authority

TABLE 4
CHARACTERISTICS OF RECORDS WITH VARIANTS

	LCNAF Found (n = 4205)		No LCNAF Found (n = 281)	
	No.	%	No.	%
Entered by OCLC Members	2,830	67.3	219	77.9
Entered as DLC/Member [†]	514	12.2	37	13.2
Entered as [blank]/Member [†]	433	10.3	19	6.8
Entered by UKM	373	8.9	0	
Entered by DLC	55	1.3	6	2.1
Encoding Level I or Blank	3,493	83.1	209	74.4
Entered after LCNAF Record Last Revised	649	15.4		

[†]LC copy entered by an OCLC member.

[†]Member input, source unknown.

records at the time a bibliographic record was created because the authority file is a constantly growing organism. In the past, new records were added at least once a week, and now they are added daily. The break between 1976 and 1977 was arbitrary, although it is perhaps logical because of a surge in retrospective conversion in the late 1970s.

It can be seen in figure 1 that a greater proportion than expected of variants that now have an LCNAF record was input before 1981, while there was a somewhat smaller proportion than expected between 1981 and August 1984, and an even smaller than expected proportion after the appearance of the revised format version of the authority file in August 1984. The proportions of variants without LCNAF records exhibit opposite relationships from the LCNAF variants. Higher proportions than expected have been input since 1981. A partial explanation for this last phenomenon appears in cases where one form was clearly in the majority before 1981, but some libraries chose to attempt to create an AACR2 form (some of which are actually now "correct") for the name after 1980. The overall observation apparent from this graph is that when authority records are available for a name, that name is less likely to be entered incorrectly into a bibliographic record.

CHARACTERISTICS OF VARIANTS

The variant headings themselves were also analyzed for the presence of certain characteristics. Table 5 shows the percentages of

variants that could be categorized as "near match" to the standard heading or to a reference, as being a single typographical error away from being an exact match to the standard heading, and as being an exact match to a reference. The largest category is "near match;" 42.9% of the records with a corresponding LCNAF record and 58.4% of the records with no LCNAF record fall into this category. As mentioned earlier, "near match" is defined as a situation in which one form is wholly contained within the other and there is no conflicting information, e.g.:

Variant and heading are a near match:

Heading: Chou, Marylin, #d 1933-

Variant: Chou, Marylin.

Heading: Edel, Leon, #d 1907-

Variant: Edel, Leon Joseph, #d 1907-

Variant and reference are a near match:

Heading: Creighton, M. #q (Mandell), #d 1843-1901.

Reference: x Creighton, Mandell, #c Bp. of London, #d 1843-1901.

Variant: Creighton, Mandell, #c bp., #d 1843-1901.

Such near matches cannot automatically be assumed to be the same person. There are many instances where they are not. For example, the heading for one sample name is **Hoffman, Peter, #d 1941-**. At the time of data collection¹² there were two entries under **Hoffman, Peter**. One was the same person; the other was not. One of the most blatant examples of this was the case of John Fletcher. The heading for the sample name is **Fletcher, John, #d 1579-1625**. At the time of data collection there were 53 entries under **Fletcher**,

TABLE 5

NUMBERS OF VARIANTS THAT REPRESENT A NEAR MATCH TO THE STANDARD OR TO REFERENCE, A SINGLE TYPOGRAPHICAL ERROR, OR AN EXACT MATCH TO A REFERENCE

	LCAF Found (n = 4,205)		No LCAF Found (n = 281)	
	No.	%	No.	%%
Near Match to the Standard	1,805	42.9	164	58.4
Single Typographical Error away from Matching the Standard	1,235	29.4	46	16.4
Near Match to a Reference	493	11.7		
Exact Match to a Reference	217	5.2		

TABLE 6
 PRESENCE OF *w ON VARIANT RECORDS

	No.	%
*w cn or dn— Name Matches an Authority Record for Another Person	135	3.0
*w 1n— Name Matches an Authority Record for Another Person	28	0.6
*w cn or dn— Form Is not Really Correct	82	1.8
*w 1n, 2n, 3n, or 4n— Form Is not Really Correct	294	6.6
*w cn or dn— Form Matches Except for Capitalization or Punctuation that Does not Affect Searching or Clustering	207	4.6
*w 1n, 2n, 3n, or 4n— Form Matches Except for Capitalization or Punctuation that Does not Affect Searching or Clustering	112	2.5
Total	858	19.1

John. Only 17 of these were the person in the sample. The others represented several other people.

Of the variants in the LCNAF group, 29.4% vary from the standard heading by only a single character. The same is true for 16.4% of the variants in the "no LCNAF" group. Single typographical errors often fall within dates, e.g.:

Heading: Fuller, Thomas, #d 1654-1734.

Variant: Fuller, Thomas, #d 1653-1734.

(These are called "typographical errors," although they might actually represent errors in handwriting before typing or might, in fact, have been deliberately created differently due to information the cataloger had in hand.) Typographical errors in dates actually do not affect searching in the OCLC system because differences in dates are ignored in the grouping process even when they are intentional. Thus, **Dickens, Charles, #d 1812-1870** and **Dickens, Charles, #d 1837-1896** are subarranged intermingled alphabetically by title of item. Most online catalogs do this also, which causes a serious problem for the searcher who is expecting high precision in addition to high recall.

Typographical errors can also appear in any part of the heading other than dates, e.g.:

Heading: Rossetti, Dante Gabriel, #d 1828-1882.

Variant: Rosetti, Dante Gabriel, #d 1828-1882.

The lack of a subfield "d" before dates was

counted as a single typographical error, even though it involves two characters, e.g.:

Heading: Brecht, Bertolt, #d 1898-1956.

Variant: Brecht, Bertolt, 1898-1956.

(These are a problem because they cause a separate alphabetical grouping of an author's works—the computer assumes the characters following "Bertolt" to be a second forename or other information indicating another person.)

Exact matches to references occur, for the most part, in situations where the LCNAF record was input into the file after OCLC's spring/summer 1987 matching project, although there are some from before this time for which an explanation has not been found. There are a few instances where exact match references have caused serious havoc in the file. At the time of data collection one blatant example was the sample name **Thompson, Charles, #d fl. 1750**. At the time of the computer matching project the authority record had a reference from **x Thompson, George** (with no dates; the authority record has since been revised and the date added to the reference). The headings on 25 records by various George Thompsons, some with very current imprint dates, were changed to **Thompson, Charles, #d fl. 1750**. Only 7 of these were correct records for this person. There was also a record for this person under **Thompson, Charles, #w cn**. The "#w cn" was there because there is an authority

record for another person under this form and the machine match "verified" this as a correct heading. As shown in table 6 this latter problem occurs with 135 (3.0%) of the variant records, and there are 28 instances (0.6%) of #w followed by "1n" attached to the variant headings in cases where the name is actually the name of another person. (The use of 1n, 2n, 3n, or 4n following #w indicates that the cataloger, rather than the machine, verified the form used in that field.)

The figures mentioned thus far from table 6 represent the situation where the name is said to match an authority record, but the record matched is for a different person. In addition there are another 695 variants with a #w that seems to indicate the form is a correct one when it is not. For 319 of these the only variation is capitalization and/or punctuation that does not affect searching, but for the remaining 376 the difference in form is significant. One can deduce from this table some difficulties in using machines for corrections. A total of 424 records (9.5% of the variant access points examined) represent situations that were left to the machine to handle with little or no human intervention, and the machine, blindly doing what it was told to do, copied or verified humans' earlier mistakes. However, it did no worse than the humans performing the same functions. They erred on 434 records (9.7%).

Again, data for the highly prolific authors in the sample exhibit a different pattern from the others, as shown in table 7.

There are proportionately many more variants that are near matches to the LCNAF heading and also exact matches to a reference in the group of names with fewer than 200 records, while the names with 200 or more records exhibit proportionately many more single typographical errors. More exact matches to references may result from more recent additions of LCNAF records to the file for the names with fewer than 200 records.

There is an exact match problem that should be mentioned but that this research, unfortunately, was not designed to count. There were a number of sample names for which there are two or more persons given the same form of name, and when there is an LCNAF record, there is a "#w cn" implying that the form is correctly used. For example, the heading **Hall, Robert D.** is in the LCNAF. This name appeared in the sample, but was not for the same person as the one in the LCNAF record. At the time of data collection there were 12 records using **Hall, Robert D.** followed by "#w cn." Three were for the same person the authority record represents. The others represented at least 5 different people. There were at least 10 of these situations that came to my attention. The reason I do not know the extent of the problem is that research assistants were not asked to examine subject content of every record if the name form was identical. Probably this should have been done, but with the high proportion of names with more than 25 records, this was

TABLE 7

NUMBERS OF VARIANTS THAT REPRESENT A NEAR MATCH TO THE STANDARD OR TO A REFERENCE, A SINGLE TYPOGRAPHICAL ERROR, OR AN EXACT MATCH TO A REFERENCE WHEN THE "LCNAF FOUND" GROUP IS DIVIDED INTO GROUPS OF NAMES WITH 200 RECORDS OR MORE OR NAMES WITH 199 RECORDS OR LESS

	200 Records or More (n = 2,666)		199 Records or Fewer (n = 1,539)	
	No.	%	No.	%
Near Match to LCNAF Record	905	33.9	900	58.5
Single Typographical Error away from Matching LCNAF Record	1,070	40.1	165	10.7
Near Match to a Reference	278	10.4	215	14.0
Exact Match to a Reference	24	0.9	193	12.5

not practical if there was to be an end to the project.

KINDS OF VARIANTS

Variants were placed into 26 categories. Many variant headings fell into more than one of these categories, and every variation found in a particular variant heading was coded. Table 8 shows the categories, the percentages of all records with variant headings ($n=4,486$) that demonstrated the particular variation, and an example of a name from each category. The total of the percentages, of course, is greater than 100% because of the large number of variant headings that demonstrated more than one type of variation. The examples in table 8 were chosen, where possible, to exhibit only the variation being illustrated. Many variants, however, have more than one variation, e.g.:

Standard form: Kabalevsky, Dmitry Borisovich, #d 1904—

Variant form: Kabalevski'i, D. B. #q (Dmitri'i Borisovich), #d 1904—1987

In tables 9, 10, and 11 the percentages of variations are grouped by broad category. Table 9 shows the distribution of the 76 variant headings from the sample records alone. Table 10 shows the distribution for the variants found during the course of the whole project. It again seemed useful to separate the highly prolific authors from the others (table 11). When this was done, there were three categories for the names with 199 or fewer records in which very large numbers of a particular error were found for just one name and greatly inflated the percentages. The results of subtracting these and then dividing by n minus the number of records representing the difference—are shown in parentheses.

It can be seen in table 9 that the category of "subfields" has a 24.5% higher percentage when there is an LCNAF record. A t -test of the two samples indicates that the difference is statistically significant ($t=2.27$). The same category in table 10 also shows a difference between the two groups that is statistically significant ($t=1.96$). This difference is likely due to the presence of subfield q , which is used to give spelled-out forenames in paren-

theses when the author predominately uses initials. This practice was begun with AACR2, and for the "no LCNAF" group, there are few times when the standard form of a name includes a subfield q . Another statistically significant difference demonstrated in table 10 is that the percentage of first forename differences is greater by 11.2% when there is no LCNAF record ($t=5.61$). (Despite the 10.9% difference for the two groups in this category in table 9, the sample sizes are too small to allow the test to show statistical significance [$t=1.13$].) Apparently first forenames vary often enough that without an authority record, consistency is more difficult to maintain. The final statistically significant difference shown in table 10 is that when LCNAF records exist, date variants exceed those for the "no LCNAF" group by 5.7% ($t=1.99$). (Again, the sample sizes in table 9 are too small to show statistical significance despite the 10.2% difference [$t=0.87$].) The difference in percentages of date variants seems to stem from the fact that famous names are more likely both to have authority records and to include dates in the authority record, and therefore there are many more opportunities for omitting or mistyping dates.

The observation about famous names and dates is verified by reference to table 11, where variations in dates account for 75.5% of the variations found for the 38 names with 200 or more records, while dates account for only 57.7% of the variations for the remaining names, a statistically significant difference of 17.8% ($t=12.03$). Table 11 shows another sizable difference between the two groups of names. Names with 199 or fewer records exhibit a much higher proportion of "Second or later forename" differences than do the more prolific names. The t -test shows this 15.7% difference also to be statistically significant ($t=12.23$). (The t -test was computed using the value in parentheses.) Like the difference in date variants, the difference in second forename variants also seems logical: catalogers and typists are more likely to be familiar with middle names of famous people, and therefore they are more likely to input the names correctly.

TABLE 8
CATEGORIES OF VARIANTS

%	Category	Examples	
		Standard Form	Variant Form
1. 68.6	Dates		
6.3	birth date or both dates included, but standard doesn't have them	Malchelosse, Gérard	Malchelosse, Gérard, #d 1896-
27.7	no dates, but standard has one or more dates	Day, Doris, #d 1924-	Day, Doris.
14.4	no death date, but standard has one	Arcaya, Pedro Manuel, #d 1874-1958	Arcaya, Pedro Manuel, #d 1874-
0.7	death date included, but is not on standard	Harper, Henry Howard, #d 1871-	Harper, Henry Howard, #d 1871-1953.
15.7	date(s) differ from standard	Marshall, Margaret, #d 1949- Dickens, Charles, #d 1812-1870.	Marshall, Margaret, #d fl. 1978- Dickens, Charles, #d 1912-1870.
2.8	no subfield d code before dates or no indicator for subfield d code	Smollett, Tobias George, #d 1721-1771.	Smollett, Tobias George, 1721-1771.
1.1	incorrect subfield code before date	Lehár, Franz, #d 1870-1948.	Lehár, Franz, #b 1870-1948.
2. 11.9	First Forename		
0.5	spelling of first forename different from standard	Grümmer, Elisabeth.	Grümmer, Elizabeth.
6.2	fullness of first forename different from standard	Pauker, Guy J.	Pauker, G. J.
0.3	first forename different name than on standard	Higginson, Alexander Henry, #d 1876-	Higginson, Henry Alexander, #d 1876-
3. 21.7	Second or Later Forename		
11.3	second or later forename or initial different form or spelling from standard	Prather, Richard S. Rossetti, Dante Cabriel, #d 1828-1882.	Prather, Richard Scott. Rossetti, Dante Bagriel, #d 1828-1882.
10.4	second or later forename or initial included/not included in variance with standard	Edel, Leon, #d 1907-	Edel, Leon Joseph, #d 1907-
4. 0.5	Forename Entries		
0.3	forename entry—spelled differently from standard	Sophocles.	Sophocoles.
0.2	forename entry—two words with comma after first, in error	Omar Khayyam.	Omar, Khayyam.

%	Category	Examples	
		Standard Form	Variant Form
5. 4.9	Surname Entries		
1.1	comma missing after surname	Ikeda, Daisaku.	Ikeda Daisaku.
3.8	spelling of surname different from standard	Stevenson, Robert Louis, #d 1850-1894	Stephenson, Robert Louis, #d 1850-1894.
6. 0.5	Completely Different Entry Word	Glareanus, Henricus, #d 1488-1563.	Loritus, Henricus, #c Glareanus.
7. 18.6	Subfields		
11.5	subfield q contents different or included/not included in variance with standard	Underwood, Francis Henry, #d 1825-1894.	Underwood, Francis H. #q (Francis Henry), #d 1825-1894.
6.2	subfield c contents different or included/not included in variance with standard	Piccolomini, Alessandro, #d 1508-1578.	Piccolomini, Alessandro, #c Archbishop, #d 1508-1578.
0.9	no or incorrect subfield code before additions other than date	Diaz, Albert James.	Diaz, Albert James, ed.
8. 14.6	Punctuation, Spacing, Diacritics, Capitalization		
0.3	includes brackets in variance with standard	Yeats, W. B.. #q (William Butler), #d 1865-1939.	Yeats, W[illiam] B[utler] #d 1865-
2.9	extra punctuation or punctuation missing, other than for comma missing after surname (already shown) [lack of a comma before #d or #e and lack of a period at the end of a heading were not counted as errors]	Huber, Miriam Blanton, #d 1889-	Huber, Miriam (Blanton) #d 1889-
0.7	necessary spaces omitted	Thirlwall, Connop, #d 1797-1875.	Thirlwall,Connop, #d 1797-1875.
5.1	diacritics different from standard	Mauriac, François, #d 1885-1970.	Mauriac, Francois, #d 1885-1970.
5.6	capitalization different from standard	FitzGerald, Edward, #d 1809-1883.	Fitzgerald, Edward, #d 1809-1883.
9. 0.04	Other	Candolle, Augustin Pyramus de, #d 1778-1841.	Candolle, #d Augustin Pyramus de, #d 1778-1841.

TABLE 9
DISTRIBUTION OF VARIANT HEADINGS BY BROAD CATEGORIES OF VARIATION:
NAMES FROM SAMPLE RECORDS ONLY

	LCNAF found (n = 50)*		No LCNAF found (n = 26)	
	No.	%	No.	%
Dates	32	64.0	14	53.8
Second or Later Forename	19	38.0	8	30.8
Subfields	18	36.0	3	11.5
First Forename	8	16.0	7	26.9
Punctuation, Etc.	5	10.0	4	15.4
Surname Entries	2	4.0	2	7.7
Forename Entries	1	2.0	0	0.0
Completely Different Entry Word	1	2.0	0	0.0
Other	0	0.0	0	0.0

*"n" represents the number of records, not the number of variations.

TABLE 10
DISTRIBUTION OF VARIANT HEADINGS BY BROAD CATEGORIES OF VARIATION:
ALL VARIANTS FOUND, GROUPED BY WHETHER LCNAF RECORD WAS FOUND

	LCNAF found (n = 4,205)*		No LCNAF found (n = 281)	
	No.	%	No.	%
Dates	2,900	69.00	178	63.3
Second or Later Forename	910	21.60	63	22.4
Subfields	793	18.90	40	14.2
First Forename	473	11.20	63	22.4
Punctuation, Etc.	604	14.40	50	17.8
Surname Entries	199	4.70	21	7.5
Forename Entries	22	0.50	0	
Completely Different Entry Word	22	0.50	0	
Other	2	0.05	0	

*"n" represents the number of records, not the number of variations.

With the exception of the categories of "Dates," "Subfields," and "First forename," the differences in the percentages of the categories in table 10 are not statistically significant. On the other hand, all differences in category percentages in table 11 are statistically significant except those for "Forename entries" and "Other." It seems clear that not only do highly prolific names dominate the database, but

they also exhibit their own patterns in terms of variants.

CONCLUSIONS

Earlier it was stated that authority records are available for all names for more than half the records in the OCLC database and that 15.1% of the sample records contain one or more names that differ from the

TABLE 11

DISTRIBUTION OF VARIANT HEADINGS BY BROAD CATEGORIES OF VARIATION:
 VARIANTS FOR NAMES FOR WHICH AN LCNAF RECORD WAS FOUND,
 DIVIDED INTO GROUPS OF NAMES WITH 200 RECORDS OR MORE
 AND NAMES WITH 199 RECORDS OR FEWER

	200 Records or More (n = 2,666)		199 Records or Fewer (n = 1,539)	
	No.	%	No.	%
Dates	2,012	75.50	888	57.7
Second or Later Forename	341	12.80	569 (387) [†]	37.0 (28.5)
Subfields	324	12.20	469 (227)	30.5 (17.5)
First Forename	183	6.90	290 (165)	18.8 (11.7)
Punctuation, Etc.	427	16.00	177	11.5
Surname Entries	148	5.60	51	3.3
Forename Entries	18	0.70	4	0.3
Completely Different Entry Word	5	0.20	17	1.1
Other	1	0.03	1	0.1

*"n" represents the number of records, not the number of variations.

[†]In each case where numbers are given in parentheses, only one name accounted for the difference between the number in the column and the number in parentheses. The percentages in parentheses are the result of dividing the numbers in parentheses by n minus the number of records representing the difference between the number in the column and the number in parentheses.

standard form given in the corresponding LCNAF. Another 5.8% of sample records have personal names that have no LCNAF record but differ from the majority of records using that name. Use of the formula for binomial proportion confidence interval¹³ on this total percentage of 20.9% shows that it can be said with 90% confidence that at least 17.7% to 24.1% of the records in the OCLC database have names in forms that may affect recall or precision in the searching of names. (This figure might be slightly higher when corporate names without authority records that differ from the standard are added.) Libraries that use these records can be fairly certain that they have similar discrepancies in their own databases, especially if they have policies of accepting cataloging as found. It is true that many libraries had their archive tapes cleaned up by a commercial service before loading them into their online catalogs. Such action would have taken care of some, but not all, of the variants identified in this paper. However, libraries can be fairly certain that they have added new discrepancies since then. After all, the

variants have continued to be added to the OCLC database, for the most part, by member libraries doing original cataloging presumably to go into their own catalogs.

However, it has also been shown that the rate of input of variant name forms has dropped as the authority file has grown. As more people become aware of the need for authority control and make more use of the LCNAF, the rate of input of variants should drop more. It is apparently not possible to keep all variants from being input while relying on humans—at least 15% of the variants were input after the most current form of heading was available in the LCNAF. This research indicates the need for the authority file to be linked to the bibliographic file in such a way that a record will have its access points checked for accuracy before it is allowed to be entered into the database.

The findings indicate that it might be possible to correct many of the existing variants with machine assistance. Between 27.4% and 29.6% of the variants vary by a single character. Many of these could be matched and corrected by machine. How-

ever, care would have to be taken to avoid collapsing under one form the records for two different people whose access points happen to differ by one character (e.g., two people with the same name born one year apart, or two people with the same name except for a one-character variation in spelling). About 5% of the variants match a reference exactly. The process of correcting these at OCLC is time-consuming and costly, but smaller databases could more easily have these corrected. Near matches could not be corrected automatically, but could be pulled out by machine for human review. It appears, then, that more than half the variants are correctable without waiting for them to be reported by someone using the records.

It should be remembered that, as happened with implementation of AACR2, in a database cleanup project the highly prolific names will receive attention first, and as they are taken care of, the proportions of what remains to be done at later stages of the project will change. This research indicates that when the names with more than 200 records have been corrected, the proportion of variants that vary by a single typographical error will drop dramatically, and the proportion of variants that are a near match to the standard will rise (see table 7).

FUTURE RESEARCH

A next step should be to study the feasibility of identifying single-character variants and near matches with computer assistance and then using human intervention to determine the ones that are indeed *true* matches. In addition, comparative studies of variation in other bibliographic databases would be useful. Further analysis of corporate names and initial analysis of geographic names, series, and uniform titles is needed.

REFERENCES AND NOTES

1. *Authority Control: the Key to Tomorrow's Catalog: Proceedings of the 1979 Library and Information Technology Association Institute*, ed. Mary W. Ghikas (Phoenix, Ariz.: Oryx, 1982).
2. For a discussion of authority control literature published in the 1980s see Arlene G.

- Taylor, "Research and Theoretical Considerations in Authority Control," *Cataloging & Classification Quarterly* 9, no.3:29-56 (1989).
3. Michael K. Buckland, *Library Services in Theory and Context*, 2d ed. (New York: Pergamon Pr., 1988), p.92.
 4. *Ibid.*, p.101, 103.
 5. Eugene Garfield, *Citation Indexing—Its Theory and Application in Science, Technology, and Humanities* (New York: Wiley, 1979); Blaise Cronin, *The Citation Process: The Role and Significance of Citations in Scientific Communication* (London: Taylor Graham, 1984).
 6. Catherine E. Pasterczyk, "Russian Transliteration Variations for Searchers," *Database* 8, no.1: 68-75 (Feb. 1985); Anne B. Piternick, "What's in a Name? Use of Names and Titles in Subject Searching," *Database* 8, no.4:22-28 (Dec. 1985); David M. Pilachowski and David Everett, "What's in a Name? Looking for People Online—Social Sciences," *Database* 8, no.3:47-65 (Aug. 1985); David M. Pilachowski and David Everett, "What's in a Name? Looking for People Online—Current Events," *Database* 9, no.2:43-50 (Apr. 1986); David Everett and David M. Pilachowski, "What's in a Name? Looking for People Online—Humanities," *Database* 9, no. 5:26-34 (Oct. 1986); Bonnie Snow, "Caduceus: People in Medicine: Searching Names Online," *Online* 10, no.5:122-27 (Sept. 1986).
 7. Elizabeth E. Fuller, "Variation in Personal Names in Works Represented in the Catalog," *Cataloging & Classification Quarterly* 9, no.3:75-95 (1989).
 8. Tamara S. Weintraub, "Personal Name Variations: Implications for Authority Control in Computerized Catalogs," *Library Resources & Technical Services* 35:217-28 (1991).
 9. Arlene G. Taylor, Margaret F. Maxwell, and Carolyn O. Frost, "Network and Vendor Authority Systems," *Library Resources & Technical Services* 29:195-205 (1985).
 10. Mark R. Watson and Arlene G. Taylor, "Implications of Current Reference Structures for Authority Work in Online Environments," *Information Technology and Libraries* 6:10-19 (1987).
 11. The fact that libraries deal with many prolific names was demonstrated by research surrounding AACR2 implementation. It was found that the amount of work necessary to resolve conflicts dropped dramatically during and after the first three years

of AACR2 implementation because the prolific authors' names showed up in the cataloging process quickly and were thus corrected first. See Arlene G. Taylor and Barbara Paff, "Looking Back: Implementation of AACR 2," *Library Quarterly* 56:272-85 (1986).

12. The caveat "at the time of data collection"

is used because a report on this research was submitted to the OCLC Office of Research in 1990, and as a result many of the examples given to illustrate variant headings in the system have been corrected by OCLC staff.

13. The formula for a 90% binomial proportion confidence interval is shown below.

$$\hat{p} - (1.65) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + (1.65) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where

p = proportion found

n = number of observations in the sample, and

1.65 = the statistic for 90% confidence.

LRTS 1991 Referees

A scholarly journal relies on the use of expert, volunteer, peer reviewers, or referees, who contribute their time, energy, and professional expertise to ensure the accuracy, relevance, timeliness, and importance of the research reported. The people whose names appear below, to whom we offer our unqualified thanks, reviewed manuscripts submitted to LRTS in 1991.—Ed.

- Marcia Bates, Graduate School of Library and Information Science, University of California, Los Angeles
- Robert Burger, University of Illinois at Urbana-Champaign
- Alice K. Chan, University of Alberta
- Mary Ellen Chijioke, Swarthmore College
- Doris Clack, School of Library and Information Science, Florida State University
- Michele Cloonan, Graduate School of Library and Information Science, University of California, Los Angeles
- Janet Crayne, University of Virginia
- Doris Dale, Department of Curriculum and Instruction, Southern Illinois University at Carbondale
- Gay N. Dannelly, Ohio State University
- Karen Markey Drabenstott, School of Information and Library Studies, University of Michigan
- Janet Gertz, Columbia University
- George E. Gibbs, University of Kansas
- Carolyn Harris, School of Library Service, Columbia University
- Janet Swan Hill, University of Colorado
- Robert P. Holley, Wayne State University
- Judith A. Hudson, State University of New York, Albany
- Ling Hwey-Jeng, Graduate School of Library and Information Science, University of California, Los Angeles
- Sheila S. Intner, Graduate School of Library and Information Science, Simmons College
- Nancy R. John, University of Illinois at Chicago
- Jay Lambrecht, University of Illinois at Chicago
- Olivia M. A. Madison, Iowa State University
- Carol Mandel, Columbia University
- Connie K. McCarthy, Duke University
- Francis L. Miksa, Graduate School of Library and Information Science, University of Texas at Austin
- Carla Montori, University of Michigan
- Miriam W. Palm, Stanford University
- A. Ralph Papakhian, Indiana University
- Marion T. Reid, California State University, San Marcos
- Margaret A. Rohdy, University of Pennsylvania
- Carlen Ruschoff, Georgetown University
- Michael T. Ryan, Stanford University
- Karen A. Schmidt, University of Illinois at Urbana-Champaign
- Brian E. Schottlaender, University of California, Los Angeles
- Charles W. Simpson, State University of New York at Stony Brook
- Arlene G. Taylor, School of Library Service, Columbia University
- Patricia M. Thomas, Stockton-San Joaquin County Public Library
- Barbara B. Tillett, University of California, San Diego
- Arnold Wajenberg, University of Illinois at Urbana-Champaign
- D. Kathryn Weintraub, University of California, Irvine
- Karen Whitney, Agua Fria Union High School, Avondale, Arizona
- Stephen E. Wiberly, University of Illinois at Chicago
- Blanche Woolls, School of Library and Information Science, University of Pittsburgh
- James B. Young, University of Pennsylvania
- Jennifer A. Younger, Ohio State University

Book Reviews

Lawrence W. S. Auld, Editor

The Bibliographic Record and Information Technology. By Ronald Hagler. 2d ed. Chicago: American Library Assn.; Ottawa: Canadian Library Assn., 1991. 331p. \$37; \$33.30 ALA member. (ISBN 0-8389-0554-4, ALA; 0-88802-261-1, CLA). LC 90-45317.

The first edition of this work, co-authored by Peter Simmons, appeared in 1982. This substantially revised, updated, and reorganized second edition treats bibliographic control in all its manifestations. This rapidly expanding field is examined from different angles, including historical, technological, and economic perspectives.

The task of writing an integrated account of modern bibliographic control necessitates working with a surfeit of difficult material. The author concedes that the new edition has "close connections to at least five current courses" (p.xiv). The resulting compression entails some paragraphs being tightly packed, complex, and too long. The book might prove overwhelming for a student or practitioner to absorb as a whole, even given the declared prerequisite experience in creating footnotes and lists of references and in using catalogs, bibliographies, indexes, and a microcomputer (e.g., for word processing). The work would perhaps serve best as background and reference reading. As in the first edition, an explanation is offered as to why a bibliography is not appended, with only certain kinds of material being cited as footnotes. This lack of a bibliography plus annotations must surely be a disadvantage for the pressured students of today.

The volume's wide-ranging scope encompasses bibliographic policies, practices, processes, tools, and utilities. Some

specific issues and topics addressed include the technical terminology of bibliographic control, functional analysis and consultation in relation to needs and changing expectations of information professionals and end users, standardization, cooperation, original and supplied records, record components, structures and formats, files, access points, search logic, devices, strategies, and contemporary automation including OPACs. A thirty-page appendix is given over to the MARC format.

A comparative approach is frequently employed. The careful, analytical writing is of very high quality, with causes of past and present practices, choices, and their consequences clearly identified and explained. Among many well-handled questions are conflicts arising between different groups of information professionals, the changing role of the Library of Congress, the functions of classification, the relationship of the catalog to other tools and services, and analytic cataloging.

This edition is organized into two main parts. Part 1 covers the principles of bibliographic control, while part 2 explores practices and standards currently followed. Inasmuch as principles are relatively stable, there is a good case for presenting part 1 as a separate publication. This would allow the author a little more space in which to do justice to his sweeping, insightful, and balanced overview. Then part 2 could be revised and issued more often, accommodating the proliferating number and kinds of applications and changes. It would serve as a current guide for planning bibliographic organization, services, and products. In the meantime, the presentation of detail and argument in one volume makes an important

contribution to professional literature.—*Alan R. Thomas, St. John's University, Jamaica, New York.*

Describing Archival Materials: The Use of the MARC AMC Format. Ed. by Richard P. Smiraglia. New York: Haworth, 1990. 238p. \$29.95 (ISBN 0-86656-916-2). LC 90-43012. Also published as *Cataloging & Classification Quarterly* 11, Nos. 3/4 (1991).

This book presents an excellent overview of archival cataloging using the MARC AMC (Archives and Manuscript Control) format. Also published as volume 11, numbers 3 and 4, of *Cataloging & Classification Quarterly*, this collection of ten articles, written by authors experienced in archival cataloging, addresses two specific audiences: the archivist whose collection needs to be cataloged for an automated system but whose cataloging knowledge is limited, and the library cataloger who desires a better understanding of archives collection management before cataloging archival materials.

In his introduction to the volume, Richard P. Smiraglia examines the similarities and differences between archival and bibliographic control. There follow four articles on the fundamentals of archival cataloging. Michael J. Fox describes the basic characteristics of archival collections, then demonstrates how the principles of descriptive cataloging apply to archival material; and Edward Swanson clearly and succinctly introduces archivists to the complex rules used to determine access points. In an article on subject control, Smiraglia discusses the usefulness of Library of Congress subject headings for archival records. Marion Matters suggests that enhanced authority records, containing biographical or historical information about the main heading in addition to the necessary cross-references, would best meet the needs of archival catalogers.

At the heart of describing archival materials is the MARC-AMC format itself. Lisa B. Weber defines the format and chronicles its history. After describing the format in some detail, she concludes her article with a brief look at the possibilities format integration holds for archival cata-

loging. Kathleen D. Roe places the MARC-AMC format within the context of automated retrieval systems designed to be shared by the library and archival communities, reminding both groups of the technological and economic constraints they face in sharing a single database.

Unlike with textual collections, the cataloging standards for archival media materials are still evolving, as becomes apparent in the final three articles of the book. Barbara Orbach identifies the cataloging tools created specifically for describing and indexing archival photographs; she also discusses issues still requiring resolution. David H. Thomas shows that cataloging archival sound recordings is similar to the cataloging of archival textual materials. Although the cataloging of archival photographs and sound recordings is well under way, the cataloging of archival maps has just begun. James Corsaro explains how the standards for archival cartographic materials are now being developed.

For those librarians and archivists who already use the MARC-AMC format, this volume is of limited value. However, for those who are just beginning to consider the cataloging of archival collections, *Describing Archival Materials* is a valuable introduction to the use of the MARC-AMC format and the principles of archival cataloging itself.—*Margaret E. Doutt, East Carolina University, Greenville, North Carolina.*

Early Bindings in Paper: A Brief History of European Hand-Made Paper-Covered Books with a Multilingual Glossary. By Michele Valerie Cloonan. Boston: Hall, 1991. 146p. (ISBN 0-8161-1971-6). LC 90-26636.

Early Bindings in Paper is the first book-length work to survey paper bindings throughout the era of hand printing. Based in part on the author's doctoral dissertation, it is really more of a handbook than a monograph. In the first part Cloonan provides a historical overview of paper bindings from the earliest known example, dated 1482, to the advent of cloth bindings around 1825. She also gives a concise but clear description of the various structures used by early binders. A detailed survey of

secondary sources, such as early book catalogs, eighteenth-century dictionaries, and pictorial images, follows. These are important because so many early paper bindings have been lost. Uses and characteristics of early paperbound books are covered briefly in subsequent chapters. Cloonan observes that paper bindings were characteristic of working libraries; they were inexpensive and allowed readers to get recent books more quickly. Their practicality for certain types of books, e.g., music, is noted also. Cloonan further argues that other scholars have erred in assuming that early paper bindings were invariably intended to be temporary, to be rebound after purchase. Many of the paper bindings proved quite durable; some received decorative treatment and were found in the libraries of such well-known collectors as Samuel Pepys.

The second part of the work is a glossary for the description of paper bindings. This is an exhaustive historical list of every pertinent term that Cloonan has encountered. While some of the terms are familiar and can be readily found in Carter's *ABC for Book Collectors* or Glaister's *Glossary of the Book*, many others are quite obscure and are found only in early sources. The main list gives English terms with definitions and foreign equivalents. Subsequent lists give German, Italian, and French binding terms. One wonders if Latin terms might not have been a worthwhile inclusion as well, since the work is historical. The lists are well organized, with numerous cross-references as well as related terms that are often treated together under a broader heading. For example, Cloonan treats all types of decorated papers in a single long entry. She includes many terms not found in the recent RBMS thesauri *Binding Terms* and *Paper Terms*. Although the scope of Cloonan's work is considerably narrower than these, it will be a useful adjunct to both with its definitions of terms and provision of foreign language equivalents.

There are a number of useful illustrations, which include both line drawings and plates. A thorough bibliography and index round out the volume.

Early Bindings in Paper offers a valu-

able resource to bibliographers, collectors, conservators, and binders. It also calls attention to a relatively neglected area and provides the groundwork for future studies. Despite the technical subject matter, the book is quite readable.—Fred W. Jenkins, *University of Dayton, Dayton, Ohio.*

Legal and Ethical Issues in Acquisitions. Ed. by Katina Strauch and Bruce Strauch. New York: Haworth, 1990. 146p. \$22.95 (ISBN 1-56024-007-5). LC 90-35841. Also published as *The Acquisitions Librarian*, no.3 (1990).

This book, a collection of fourteen papers authored by twenty individuals, was also published as number 3, 1990, of *The Acquisitions Librarian*. The introduction is written by the editors, not by Bill Katz as stated in the publication announcement.

William Hannay's essay on antitrust issues in publishing describes two problem areas. Regarding mergers and acquisitions, Hannay stresses that it is the protection of competition that motivates antitrust laws—not the protection of freedom of expression. However, he does not elaborate on the concern about media concentration and its impact on freedom of expression. The other issue discussed is the discriminate pricing policy of some publishers that favors the large bookstore chains. Suzanne Krebsbach's paper, "Acquisitions and the FTC: A Brief Introduction," is indeed brief (five pages) and cites the same case on price discrimination discussed in the previous paper by Hannay. In another brief contribution, Margaret Axtmann discusses legal and ethical issues related to publisher advertising. She focuses on the guidelines established by two works, "Guides for the Law Book Industry" and ANSI Standard Z39.13. In a similar vein Marcie Kingsley and Philip Berwick provide a practical perspective on billing problems that acquisitions librarians encounter. The substance of these papers relates more to poor service and poor customer relations than to legal or ethical issues.

A couple of papers deal with gifts to libraries, but neither deals in a substantive way with some of the real dilemmas in

which librarians find themselves as the recipients of charitable contributions. Three nonlibrarians authored the paper "Gifts—The Answer to a Problem." The paper demonstrates little understanding of library operations, and the suggestion that gifts can offset the loss of financial resources is simply uninformed. Joyce Ogburn discusses the legalities of acquiring software and the complexities associated with this format. Meta Nissley provides a well-written article on newer technologies and their licensing agreements. Her paper includes a realistic discussion of the burden of enforcement and the need to negotiate changes in such agreements to meet the needs of the individual institution.

In "Journeyman of the Printing Office," Suzanne Freeman and Barbara Winters provide a wide-ranging essay on ethical and fiscal problems associated with the involvement of librarians in both the acquisitions and authorship processes. They refer to research that suggests that any article published is of interest to only 10 percent of the people in a given discipline, that only 10 percent of the literature of a discipline is "real scholarship," and that few papers have an impact on subsequent scholarship. They point to problems with the peer review process. They also point to the ethical and fiscal dilemma presented to librarians by the fact that *Chemical Abstracts* increased its journal titles covered by more than 1,000 percent between 1927 and 1976 and yet experienced a corresponding decrease in the percentage of the total literature covered. To their credit, Freeman and Winters see some solutions. The shift to consumer-oriented knowledge may relieve pressure on libraries on the output side, and the refusal of librarians to author articles for nonreviewed journals might help on the input side.

James Coffey has contributed a paper about the expressed and implied provisions of contracts with vendors. He feels that legal issues are rarely encountered. His emphasis on competence as an ethical obligation is on target. Other papers deal with obscenity and juveniles, claiming periodicals, negotiating service charges, and discard policies. Rosann Bazirjian's paper on discard policies makes an interesting

attempt to apply philosophical theories to the ethical issues in weeding. She also comments on the ethical implications of using much-needed space for outdated editions and extra copies of unused books.

The book has no index and the last date for references cited is 1988. By and large, no new observations are offered in these papers. A few provide points of departure for other investigations.—Don Lanier, *University of Illinois at Chicago, Library of Health Sciences, Rockford, Illinois.*

Library Cooperation and Networks: A Basic Reader. By Anne Woodsworth with the assistance of Thomas B. Wall. New York: Neal-Schuman, 1991. 208p. paper, \$39.95 (ISBN 1-55570-088-8). LC 90-28521.

The author's stated objectives, to write a basic reader on library cooperative efforts and to provide a textbook on this subject, are both successfully accomplished. The book brings together what until now has been scattered in the literature.

The text begins with an overview of networking and identifies the current directories and literature available. A brief history of library cooperation includes private-sector endeavors. In discussing different cooperative structures, the author includes organization by purpose, type of library, geographic boundaries, and political jurisdiction. The perspectives of academic, public, special, and school libraries as well as archives are carefully presented, including their reasons, their expectations, and their levels of participation. Specific cooperative projects are presented, which include programming, resource sharing, bulk purchases, staff training, and communication and information dissemination.

Diagrams of governing structures are provided with the legal issues addressed. Sources of funds for library networks are discussed along with issues of fees to member libraries as well as to end users. Management issues, planning and decision making, contracts, and ownership of shared resources are addressed. The roles of national libraries (in Canada as well as in the United States) and of state governing bodies are identified as strong influences in network development.

Although library cooperation is not dependent on the use of technology, computers have greatly influenced the activities and services of networks, and the issues concerning interconnecting computer networks are discussed. NREN is included in these discussions.

One of the barriers to cooperation identified in 1978 by the National Commission on Library and Information Science was attitude.¹ The chapter on behavioral issues discusses the personal feelings and concerns of library staff and organization, including the fears of technology. The concluding chapter identifies failed efforts and lists fourteen "real barriers" to library networking.

Each chapter is followed by several questions and issues to stimulate discussion and further study as well as references and additional readings. This open-ended style leads readers toward present as well as future issues. Additional materials include a twelve-page bibliography, a list of acronyms, a glossary, sample bylaws for cooperatives and consortiums, membership agreements, and a membership survey.

From the 1917 interlibrary loan code to NREN, this book covers the formation, development, and future of library cooperation. It is recommended for anyone, including those outside the profession, seeking answers to problems and concerns, and guidance for future cooperation.—*Diane D. Kester, East Carolina University, Greenville, North Carolina.*

REFERENCE

1. National Commission on Libraries and Information Science, Task Force on the Role of the School Library Media Program in the National Program, *The Role of the School Library Media Program in Networking* (Washington, D.C.: NCLIS, 1978).

Library of Congress Subject Headings: Philosophy, Practice, and Prospects. By William E. Studwell. Supplement no.2 to *Cataloging & Classification Quarterly*. New York: Haworth, 1990. 120p. \$22.95 (ISBN 1-56024-003-2). LC 89-26970.

Studwell's aim in this slim volume is to offer some theoretical principles for the

improvement of the Library of Congress subject heading system, or at least to set up a platform on which a set of principles can be developed. This he has accomplished.

The author has divided his suggestions into those concerning the structure of LC subject headings, the terminology used, the specificity of the list, and the presentation of data. Many of the structural principles offered will sound familiar to anyone who has given thought to the problems of subject presentation: the elements used in headings should be simple, consistent, and have logical relationships; inversions should be avoided whenever possible; topics should have preference over place as the initial element in a heading. Studwell does, however, have some suggestions upon which there may be much less agreement among subject catalogers: headings should be "diagrammable"; "rival headings" such as ART, AMERICAN and ART—UNITED STATES should be eliminated, as should such "reverse patterns" as SATELLITES—MARS and MARS (PLANET)—ATMOSPHERE; and new visual cues such as the use of slashes and equals signs should be considered.

The suggestions regarding the terminology of subject headings seem equally familiar: headings should use natural language, be consistent and clear, and be sensitive to social issues; the terms used should successfully differentiate disciplines and topics; names used in subject headings must be consistent in form with the same names used as main or added entries, except for those of governments; and there should be generally accepted guidelines for the formation of headings that are not listed in LC publications.

With regard to specificity of headings, Studwell's proposals seem intended to bring about a level of standardization that would comfort a great many catalogers faced with original cataloging problems: the detail in headings must be evenly developed; full-period subdivisions should be created for geographical areas; and standard-period subdivisions should be established for use under various types of headings.

Under presentation of data, the suggestions are intended to create a subject head-

ing list that would be more "cataloger-friendly" and result in a more user-friendly catalog. Data should be presented more clearly, more completely, and more concisely; there should be readily available lists of subdivisions; headings should be self-explanatory, and there should be no subdivisions that can be applied only under certain conditions. More guidance should be provided including instructions for using complementary headings; a comprehensive listing of "gathering levels"; provision for handling subjects that reflect a main or added entry; and a determination of when to break off using juvenile subjects.

A far smaller section on application deals with the proposed use of secondary subject headings, the number of subjects to be used, and the order in which they are to be presented. A final, brief discussion entitled "The Future" makes it clear that while Studwell recognizes that the degree of automation of the catalog will affect the ways in which subjects can be approached, his principles are to be considered applicable to either manual or online catalogs.

The author's hope that his suggestions will lead to "the development of a comprehensive and widely agreed upon set of principles" (p.11) by the end of the century seems perhaps optimistic, as does his belief that the suggestions made will create a subject catalog that is easier for the user. It is hard to imagine the catalog user who will not blink when faced with a heading such as CATS/ART or WOMEN—LAWYERS, as well as with one of the "logical strings" in current use such as EDUCATION—FRANCE—PARIS—HISTORY—BIBLIOGRAPHY—CATALOGS.

The book seems peculiarly limited in that Studwell makes little or no acknowledgement of the contributions of history in the development of subject headings; no work of Charles Cutter appears even in the general bibliography, which is described by the author as "good background material on the subject" (p.115). For this reviewer, there was also a continued mental grating caused by the author's inability to decide whether "LC" is a singular or plural noun.

From reading the suggestions in this

volume, one can easily envision a time in the not-too-distant future when all approaches to the subject catalog may involve an intermediary. For that reason alone, this book should be read by reference librarians as well as subject catalogers, and by public librarians as well as by the academic and research types who will be less uneasy with the world it describes.—*Constance Rinehart, University of Michigan, Ann Arbor.*

Library Technical Services: Operations and Management. Ed. by Irene P. Godden. 2d ed. San Diego: Academic Pr., 1991. 284p. \$49.95 (ISBN 0-12-287041-7). LC 90-25393.

This second edition of *Library Technical Services* updates a standard work. An introduction by Irene P. Godden defines the boundaries of technical services; individual chapters cover technical services administration, library automation, acquisitions, bibliographic control, and preservation. These detailed essays pull together both foundation and recent literature, with a focus on the problems of large libraries. As noted in the preface, in order to reflect significant changes, discussion of circulation has been dropped in favor of expanded treatment of the role of automation and networking.

Because many technical services librarians are also managers, the chapter on administration by Leslie A. Manning is complemented by sections on management in the other chapters. Manning discusses management concepts specific to the role of technical services units in larger organizations. An emphasis on strategic planning and personnel management will guide managers in the increasing delegation of technical services tasks to nonprofessional staff.

Like management concerns, automation affects every aspect of technical services operations today and will continue to do so in the future. Karen Horny's chapter on the "ideal and reality" of library automation describes the integrated system ideal and summarizes the history of technical services automation, with attention to cooperation and networks, formats, and standards for resource sharing. Service im-

provements continue to blur traditional boundaries between public and technical services. The bibliography is brief, recognizing the quick pace of revision in the field.

The chapter on acquisitions by Sara C. Heitshu focuses on ordering and receipt of monographs in large libraries, though some attention is given to serials, as well as to the needs of smaller libraries acting either alone or in consortia. Heitshu covers vendor relations and compares manual and automated file systems for acquisitions record keeping. The access/ownership debate is covered in a section on "alternatives to ownership."

Bibliographic control, as Betty G. Bengtson notes, is in transition today, as libraries move from card to online catalogs. Bengtson's discussion of the theory and history of descriptive cataloging, subject cataloging, and classification provides sufficient context for the extended discussion of MARC, automated cataloging, and retrospective conversion. The impact of standardization and cooperation on department organization and management issues, as well as interrelationships within the library, is also noted.

A. Dean Larsen and Randy H. Silverman identify the essential preservation concerns of environmental control, house-keeping, and binding. Sections on fire and water disasters are complemented by a section on insurance, a topic that might be overlooked by preservation librarians. The need for preservation education is stressed throughout. A concluding glossary (unfortunately marred by typographical errors) defines preservation treatment options and compares relative costs.

Though the comprehensiveness and currency of the information make this a likely textbook, its high cost and availability only in hardcover format may limit its assignment. However, as a text this title is preferable to *Technical Services Today and Tomorrow*, edited by Michael Gorman (Libraries Unlimited, 1990) because the authors provide more explanation of basic concepts and more comprehensive bibliographies. The information is both more scholarly and more up to date than that found in Donald L. Foster's *Managing the*

025.431092*

CLASSIFY YOURSELF

as a Dewey expert.

These powerful aids
will help you.

IN CELEBRATION OF REVISED 780:
MUSIC IN THE DEWEY DECIMAL
CLASSIFICATION EDITION 20. 1990.
Paper. \$20.00

In-depth assessment of revised 780
music class and suggestions for suc-
cessful application.

CLASSIFICATION THEORY IN THE
COMPUTER AGE: CONVERSATIONS
ACROSS THE DISCIPLINES. 1989.
Paper. \$20.00

Essays on classification research;
classification and computer science;
classification theory and practice.

SUMMARIES. Booklet contains the 10
main classes, 100 divisions and 1000
sections of the Dewey Decimal
Classification. Single copy: \$3.00

POSTERS. 20" x 28" contemporary
poster honoring Melvil Dewey. Single
copy: \$5.00

BOOKMARKS. Cartoon figure studies
the 10 main classes of DDC. Double-
sided. Package of fifty: \$5.00

OCLC  **Forest
Press**

To order, write Forest Press OCLC,
6565 Frantz Road, Dublin, Ohio 43017-
3395. National 800-848-5878. Ohio
800-848-8286. Canada 800-533-8201.

*classifier

Catalog Department (Scarecrow, 1987).—*Ellen Crosby, University of South Carolina, Columbia.*

Manheimer's Cataloging and Classification: A Workbook. 3d ed. Rev. and expanded by Jerry D. Saye with Desretta V. McAllister-Harper. New York: Marcel Dekker, Inc., 1991. 274p. paper, \$29.75 (ISBN 0-8247-8493-6).

Jerry Saye, associate professor in the School of Library and Information Science at the University of North Carolina at Chapel Hill, was asked by Martha Manheimer to prepare this third edition of her classic workbook, which was last revised in 1980. He and Desretta McAllister-Harper have remained faithful to the structure of the previous edition, while adding one-third more card examples, expanded explanations of classification schedules and subject headings, and more detailed answers to the exercises.

The first five chapters consist of 240 card examples illustrating selected rules from the *Anglo-American Cataloguing Rules*, second edition, 1988 revision (AACR2R) for description of monographs, choice of access points, and establishment of headings for persons, corporate bodies, and uniform titles. Many of the card examples have been revised or replaced by new examples that better illustrate the rule or describe books of current interest; these have been typeset for easy reading. The remaining three chapters cover *Dewey Decimal Classification*, based on edition 20; Library of Congress Classification schedules as of 1989; and *Library of Congress Subject Headings (LCSH)*, 12th edition, with reference to the *Library of Congress Subject Cataloging Manual*.

Because this is a workbook, the user will need to have access to AACR2R, the

previously mentioned classification schedules, and LCSH in order to complete the exercises and gain full benefit from the text. The workbook is designed to be used in cataloging and classification classes by providing a quick overview, in AACR2R order, of the major rules for cataloging monographs, using well-chosen and interesting examples. Textual information is brief but valuable, with the intention that an instructor will expand on the details as needed. The workbook is also an excellent source of exercises on classification and subject headings, and provides clear, detailed answers. However, it includes only six exercises showing title pages to be used to practice descriptive cataloging.

The two weaknesses of the workbook are its lack of title page illustrations to show how cataloging was derived and its exclusive use of catalog card examples with no mention of the MARC format. The authors have attempted to remedy this by offering at cost sets of transparency masters (not seen by this reviewer) with either supporting title page illustrations or examples in the MARC format. If a workbook is desired that emphasizes preparation of MARC tag workforms from title pages, a better choice is *Cataloging Books: A Workbook of Examples*, by William Studwell and David Loertscher (Libraries Unlimited, 1989).

This third edition of Manheimer's workbook provides an excellent summary of traditional cataloging and classification practices. An instructor could be very creative in building lectures and assignments around it, but the brevity of its information limits its use as a self-study tool. It fulfills its purpose as a supplement rather than a substitute for more comprehensive texts on cataloging theory.—*Lori Osmus, Iowa State University, Ames.*

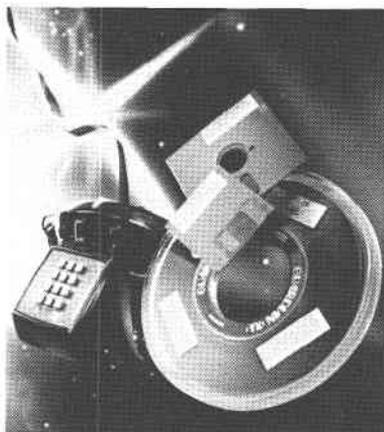
Something NEW from MARCIVE, Inc.!

Introducing
MeSH Authorities Processing

MARCIVE replaces obsolete MeSH headings with the latest MeSH headings and provides matching, deblinded NLM MARC authority records for loading into local systems.

At the same time, LC names, titles, and subjects can be examined and upgraded at no additional cost.

And all this is done with fast turn-around time, high quality, and low prices.



Please call or write for more information about how MeSH processing can benefit your library, consortium, patrons, and you:

Marcive®

P.O. Box 47508

San Antonio, TX 78265

1-800-531-7678

(512) 646-6161

FAX (512) 646-0167

Letter

From F. W. Lancaster, Professor, Graduate School of Library and Information Science, University of Illinois, Urbana, IL 61801

In his otherwise rather favorable review of my book *Indexing and Abstracting in Theory and Practice* (LRTS, January 1992), [Hans H.] Wellisch criticizes me for in-

cluding references to sources in Czech, Russian, and Danish (it was actually Norwegian). I find this rather ironic in view of the fact that, in the past, he has condemned the *Annual Review of Information Science and Technology* for tending to review only sources in English (*International Library Review*, 1973, 5, p.161).

INDEX TO ADVERTISERS

ALA	255
Baker & Taylor	130
Brodart	cover 2
EBSCO	133
Haworth	129
Library Technologies	cover 3
Marcive	251
OCLC	249, 254
Roth Publications	cover 4
H. W. Wilson	134
Winnebago	256

Instructions for Authors

Please follow these procedures for manuscripts to be submitted to *Library Resources & Technical Services*:

1. Submit original, unpublished manuscripts only. Do not submit manuscripts that are being considered for publication in other venues. Articles that advance knowledge in collection management and development, acquisitions, and technical services are preferred. (For further information see "Editorial Policy" in *Library Resources & Technical Services* 35: 357 (1991).)
2. Write the article in a grammatically correct, simple, readable style. Whenever possible avoid jargon and acronyms. For spelling and usage consult the latest edition of *Webster's Ninth New Collegiate Dictionary*, supplemented by the latest edition of *Webster's Third New International Dictionary of the English Language*; prefer the first spelling. Verify the spelling and accuracy of all names in an appropriate reference. Consult *The Chicago Manual of Style*, 13th ed., revised and expanded (Chicago: Univ. of Chicago Pr., 1982) for capitalization, abbreviations, usage of numbers, etc.
3. Give the article a brief title; if the title does not fully describe the content of the article add a brief subtitle. On the first page of the manuscript give the article title, the name(s) of the author(s), and the position title, institutional affiliation, and address of each author.
4. On the second page of the manuscript give the title and subtitle (if any), followed by a brief, informative abstract, typed double-spaced. Do not identify the author(s) here or elsewhere in the manuscript.
5. Type the manuscript, double-spaced, on 8 1/2-by-11-inch nonerasable paper. Use fresh, bright typewriter or computer ribbons or cartridges. PLEASE TYPE EVERYTHING DOUBLE-SPACED.
6. Follow the examples and suggestions in chapter 12 of *The Chicago Manual of Style* in designing tables. Submit each table on a separate page at the end of the manuscript. Indicate the preferred placement in the text with an instruction in square brackets. Provide each table with a brief, meaningful caption. TYPE TABLES DOUBLE-SPACED THROUGHOUT.
7. Be prepared to supply camera-ready copy for all illustrations. Accompany the manuscript with a photocopy of each, and a brief, meaningful caption noted on the verso.
8. Submit all notes and references on separate pages at the end of the text, preceding any tables or illustrations. PLEASE TYPE ALL NOTES AND REFERENCES DOUBLE-SPACED. Use superscript numbers throughout the text, but do not type the numbers as superscripts in the notes and references, and do not indent the first line. Use references to document the text, not to amplify it. Note that a shortened form (not *op. cit.* or *loc. cit.*) is used for subsequent references to a previously cited work. If no other reference intervenes, use "Ibid." to take the place of the elements of the previous reference that apply. Do not underline "Ibid." VERIFY EACH CITATION VERY CAREFULLY. Use chapter 17 of *The*

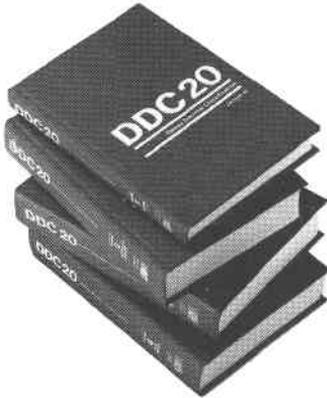
Chicago Manual of Style in formulating bibliographic references.

9. Send three complete copies of your manuscript, including illustrative material, to: Richard P. Smiraglia, Editor, Library Resources & Technical Services, School of Library Service, Columbia University, 516 Butler Library, 535 W. 114th St., New York, NY 10027.

In general, the LRTS editorial staff follows the *Guidelines for Authors, Editors, and Publishers of Literature in the Library and Information Field*, adopted by the American Library Association Council in 1983 and available from the ALA Executive Offices. Information about copyright policies also is available from ALA headquarters.

Time to order your new Dewey.

Expanded to four volumes, up-to-date, the Dewey Decimal Classification organizes today's information with current topics and terms.



DDC 20, now in its third printing.

New features:

- a manual to guide the classifier
- a revised index for easier subject access
- more instruction notes
- more summary schedules for quick subject overview

Make your world a little more orderly, and order today. Dewey Decimal Classification and Relative Index, Edition 20. 4 volumes, printed on permanent paper. ISBN 0-910608-37-7. \$225.00. Send your order today to Forest Press OCLC, 6565 Frantz Road, Dublin, OH 43017-3395.

OCLC  **Forest Press**

Publisher of the Dewey Decimal Classification®
A division of OCLC Online Computer Library Center, Inc.



SIMPLE SOLUTIONS TO YOUR AUTOMATION CHALLENGES

from ALA Books and the Association for Library Collections and Technical Services

New!

202(+) Software Packages to Use in Your Library: Descriptions, Evaluations, and Practical Advice

101 Micro Series

Patrick R. Dewey

The second edition of this versatile reference tool provides the reader with up-to-date, user-friendly information on more than 250 recommended software packages. These comprise library skills, CD-ROM disks, accounting, online catalog programs, and other specialized categories. Vendor, price, hardware requirements, uses in the library, capacity, description, nature and quality of documentation, similar or related programs, and additional sources of information are provided.

\$27.50pbk. 190p. 1992

ALA Order Code 0582-X-0011

Preservation Microfilming: Planning & Production

*Papers from the RTSD Preservation
Microfilming Institute, New Haven, CT,
April 21-23, 1988.*

Gay Walker, editor

\$12.00pbk. 72p. 1989

ALA Order Code 7324-8-0011

Automating the Small Library

*LAMA Small Libraries Publications
Series, #18*

William Saffady

William Saffady, author of *Introduction to Automation for Librarians (ALA 1989)*, brings the reader a much needed overview of computer applications appropriate to small libraries. He focuses his discussion on six areas of application: circulation control, acquisitions and serials control, descriptive cataloging, reference services, online catalogs, and administration.

\$5.00pbk. 16p. 1991

ALA Order Code 5745-5-0011

A Core Collection in Preservation

Lisa L. Fox

"An extremely useful and inexpensive bibliography [listing] works on applications of micrographics for library preservation as part of a holistic approach to the national preservation effort."

—*Microfilm Review Quarterly*

\$5.00pbk. 15p. 1988

ALA Order Code 7224-1-0011

ALA BOOKS

50 East Huron Street Chicago, IL 60611 1-800-545-2433; press 7 to order

Manage students...



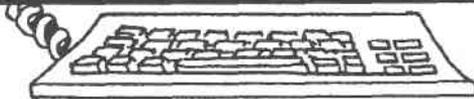
Manage materials...



District-wide locations...



Manage, search, and find with a few keys.



And at any step get the Customer Support you need. Fast.

With our IBM/compatible or Macintosh CIRC software, you can easily manage up to 100,000 patrons. With a few keystrokes, check in and out, renew and reserve, calculate fines, review checked-out and overdue materials and more.

With our IBM/compatible or Macintosh CAT, patrons can instantly look up materials simply by entering key words, phrases, subject, author, title or notes. And that's even with approximate spellings.

Either the CIRC or CAT software stores up to 300,000 materials. And both work together creating a single, versatile database for easier materials management.

To track and manage materials in all libraries in your district, simply take

command at your IBM or compatible unit with our UNION CAT.

If you ever need help to keep your library running smoothly, we guarantee Customer Support call back in two hours or less!

CIRC, CAT, and UNION CAT--just a few of the many ways we can help you make better use of your time. Call Winnebago Software today for more information.

1-800-533-5430, ext A-3



**Winnebago
Software
Company**

457 East South Street, P.O.Box 430
Caledonia MN 55921



Authority Control

- LC MARC authority files - names, subjects, titles, (updated weekly)
- Manual review of unlinked headings by professional librarians
- Deblinded LC authority records written to magnetic tape
- Inexpensive machine only processing option available
- Update service with on-going notification of changes
- Full service program, including deduping, item field builds, smart barcoding

Before you select an authority control vendor, ask what percentage of your library's headings are likely to be validated against LC authority records. Then call LTI.

"A Commitment to Quality"



LIBRARY TECHNOLOGIES, INC.

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

*Available
Now!*

The largest, most comprehensive, and most current poetry index ever created with the power, speed, and simplicity of CD-ROM.

A unique application of the latest technology to the humanities, **POEM FINDER on Disc** will enable your library to locate hundreds of thousands of poems.

The logo for POEM FINDER on Disc. The words "POEM FINDER" are in a large, bold, black, sans-serif font. "on Disc" is written in a smaller, black, cursive script font below it. The entire text is enclosed within a double-lined circular border.

- **The most comprehensive** - Indexes over 270,000 poems. Almost four times the size of any other poetry index. Coverage runs from antiquity to today. Provides access to poetry published in anthologies, single-author collections, and periodicals.
- **The most current** - Indexes all the traditional and the most contemporary poetry. Accesses over 80,000 of the most current poems from 1985 to date. Guaranteed to be cumulatively updated every two years.
- **Keyword searching through all fields** - You can search by author, translator, poem title, book or periodical title, first line. Complete Boolean searching with up to four operators (And, Or, Not, Between). Every citation provides full bibliographic information for the poem source including page numbers.
- **Extremely user-friendly** - POEM FINDER employs the Sony® Corporation's Questar™ interface, making it ideal for patron use.
- **The most affordable** - With a database that's four times the size and a price that's less than half of any other electronic poetry index, your library can't afford to be without POEM FINDER.
- **Full text on demand** - For poems not under copyright, our unique Poems by PhoneSM Service will provide your library with the full texts, by fax or mail, within 24 hours. The full texts of over 100,000 poems are available through Poems by Phone.
- **Fully networkable** - Unlike other electronic poetry databases, POEM FINDER can be licensed for use on LANs. Call us for details.
- Available on 60 day approval

To order call k



185 Great Neck Road
Great Neck, NY 11021
(800) 327-0295
Fax (516) 829-7746

POEM FINDER on Disc and Poems by Phone are trademarks and service marks of Roth Publishing, Inc. All rights reserved. Sony is a registered trademark and Questar is a trademark of Sony Corporation. All rights reserved.
Computer requirements: PC/MS-DOS version 3.0 or higher, 640K primary memory (RAM), CD-ROM drive using Microsoft extensions 2.0 or higher.