Volume 7
Number 2

ISSN 0730-9295
Annos XX Natus

Information Technology and Libraries

June 1988

CONTENTS

ITLBDC 7(2) 105–216 (1988); Inf Technol Libr


DINOSAURS and Other Archosaurs

Annos XX Natus

Information Technology and Libraries

Volume 7, Number 2: June 1988

CONTENTS

ITLBDC 7(2) 105–216 (1988); Inf Technol Libr

Information Technology and Libraries

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

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 103 of the March 1988 issue.

Manuscripts of articles, communications, and news items should be addressed to the editor: William Gray Potter, *Information Technology and Libraries*, Hayden Library, Arizona State University, Tempe, AZ 85287. Copies of books for review should be addressed to: Karin A. Trainer, *ITAL Book Reviews*, Yale University Library, 120 High St., P.O. Box 1603A, Yale Station, New Haven, CT 06520. Copies of software for review should be addressed to: David T. Buxton, *ITAL Software Reviews*, A349 Main Library, University of Arizona, Tucson, AZ 85721. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $17.50, is included in membership dues. Nonmembers may subscribe for $35 per year in U.S., Canada, Mexico, Spain, and other PUAS countries; $40 in other foreign countries. Single copies, $9. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. *Postmaster:* Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Circulation and Production: Central Production Unit (Eileen Mahoney, Director of Production and Distribution Services; Chere Elliott, CPU Supervisor; Sara Ringle Drummond, Debra M. Hall, Eloise L. Kinney, Paula Sedor, and Debra J. Teague), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk*, and *Science Abstracts Publications*. Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social & Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, NewSearch*, and *Social Sciences Citation Index*. Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

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

Guest Editorial: Happy Birthday!

A dog-eared, rather faded magazine on my desk bears the title *Journal of Library Automation*. It is volume 1, number 1, and I've just dug it out to confirm a recent observation by the current *ITAL* editor that seemed incredible. Yes, the date is March 1968, so it must be true: *ITAL* and its predecessor are now twenty years old.

The *Journal* had its roots, of course, in even older history: the formation of the Information Science and Automation Division (now the Library and Information Technology Association). On January 27, 1966, acting on a petition with some 800 signatures (after extended debate), the American Library Association voted to establish its fourteenth division, with "the development and application of electronic data processing techniques and the use of automated systems in all areas of library work" as its area of concern. Before this date, no unit held broad responsibility for library automation, and there was no effective way for librarians involved with automation to communicate or learn from each other's experience and no way for the national professional association to provide leadership to those in need of information and guidance. One of the most important objectives of the division, therefore, was "the provision of a forum for discussion of library automation problems and experiences, and other means of communicating information." And high on the list of specific activities suggested by the division's founders was "the establishment of a journal which would pull together articles on library automation."

The next year or two were busy but happy ones, as those of us involved set out to put together the organizational pieces. One spring day in particular stands out in memory. I had flown to Chicago on a morning flight to meet with headquarters personnel about the details of publishing the journal. A particularly vigorous cold front had moved through the area, so the trip meant a delayed flight and walking at an acute angle with the wind blowing my briefcase, stuffed with ALA documents, straight out behind me. Getting home from Chicago to Washington, D.C., was even worse, involving six airports and almost fifteen hours on a plane. But no matter: Fred Kilgour, who had helped hugely in getting the division started, had also agreed to take on the task of editor, and the journal was finally on its way.

Since then the division has grown and changed in a number of ways, as one would expect. Some ideas, such as a bank of computer programs for use by other libraries planning similar applications, a clearinghouse for information on library automation projects, and written evaluations of library automation equipment, proved impractical or never took hold. Other ideas, however—tutorial seminars, preconference meetings and other educational programs, attention to standardization, and distribution of bibliographic data in machine-readable form—have succeeded beyond the founders' dreams.

The division has also shown an important and impressive ability to renew itself in order to meet new needs. In 1973, acting on the recommendations of a special Committee on Objectives, the division added new sections with responsibility for audiovisual and cable technology and changed its name (as well as the name of the journal) to reflect the broader scope. In 1985, reflecting even broader and more diffuse interests, the division again reorganized, this time into some thirteen different interest groups and sixteen different committees.

These changes are in part a reflection of developments in the field of library automation itself. Gone are most of the "home brewed" systems that were so exciting and challenging to build, replaced by commercial systems that are much more sophisticated and much less expensive, even if no less traumatic to implement. The number, size, and importance of online databases has grown dramatically. And there has been a gradual, encouraging increase in cooperative, networked, and consortial automation activities. These trends seem likely to continue; they have made the field more stable and mature, and they will continue to make it exciting as well.

As it was expected to do, the journal has documented and reflected these developments. Here in the first issue are articles on three locally developed systems: an acquisitions system at Texas A&I, a book catalog at Stanford, a fund accounting system at Brown (that article written, I note with interest, by a young systems librarian named Robert Wedgeworth). By 1975, Richard De Gennaro was heralding the "second decade" of library automation with many of the developments mentioned above, and by 1983, Susan Martin was citing new technologies, such as satellites and lasers, and the increase in standards and cooperative activities as signs of a "growing maturity" (while worrying about the disintegrating influence of microcomputers).

For the journal's success during the past twenty years, much of the credit must go to the succession of outstanding editors with which it has been blessed. Following Fred Kilgour were Sue Martin, Bill Mathews, Brian Aveney, and (for the last five years) Bill Potter. With their able leadership, and the help of a conscientious and knowledgeable editorial board, the journal has amply fulfilled its original objectives and will undoubtedly continue to do so.

STEPHEN R. SALMON

Editor's Note: Stephen R. Salmon, now Chairman of Carlyle Systems, Inc., was the founding president of the Information Science and Automation Division, the predecessor of the Library and Information Technology Association.

When You Need an Answer, and You Need It Fast, and It Has to Be Right—Turn to

WILSONLINE®

ONLINE ACCESS TO THE WILSON INDEXES

"Recommended for all libraries." —REFERENCE BOOKS BULLETIN

"A sophisticated search system that will impress searchers accustomed to the most advanced techniques." —DATABASE

WILSONLINE—Recipient of the 1985 Dartmouth Medal

Current Bibliographic Data

Indexing everything from popular magazines to professional journals, WILSONLINE covers

- More than half a million articles in 3,500 periodicals each year
- 80,000 books annually
- 24,000 government publications annually
- More than 600,000 new records each year
- A total of more than 3 million records to date.

Quality

Updated at least twice each week, WILSONLINE databases contain the same dependable information as the familiar printed versions.

Affordability

WILSONLINE's pricing makes it affordable for more libraries than any comparable service. WILSONLINE becomes even more affordable when you add up the fees you don't have to pay: No start-up fee...No charge for saved searches...No royalty charges added to connect-time rates.

User-Friendly WILSONLINE Features

• Up to 43 different access points to bibliographic records • Searching of up to 8 databases simultaneously • Proximity searching • Nested Boolean logic • Automatic substitution of preferred form of subject headings and personal and corporate names • Free text searching • Controlled vocabulary searching • A toll-free number for user assistance • Online help command and explain messages.

WILSONLINE Databases

• Applied Science & Technology Index • Art Index • Bibliographic Index • Biography Index • Biological & Agricultural Index • Book Review Digest • Business Periodicals Index • Cumulative Book Index • Education Index • Essay and General Literature Index • General Science Index • GPO Monthly Catalog • Humanities Index • Index to Legal Periodicals • Index to U.S. Government Periodicals • LC/Foreign MARC File • LC/MARC File • Library Literature • MLA International Bibliography • Readers' Guide Abstracts • Readers' Guide to Periodical Literature • Social Sciences Index • Vertical File Index • Wilson Journal Directory • Wilson Name Authority File • Wilson Publishers Directory

To Open Your WILSONLINE Account Call Toll-Free: 1-800-367-6770

In New York State, call 1-800-462-6060; in Canada, call collect 212-588-8400.

THE H.W. WILSON COMPANY
950 University Avenue
Bronx, New York 10452
(212) 588-8400

System-System Interaction in Computerized Indexing of Visual Materials: A Selected Review

Diane Brooks

International computerized indexing systems for visual materials are discussed. The systems included were among the best in the literature for their time or have significant potential and represent the state of the art in their respective countries. Technologies available for reproducing images are also investigated, highlighting videotex, videodisc, and digital storage. The interfacing of these technologies with each other and with online, postcoordinate retrieval systems could facilitate the ultimate achievement in visual indexing.

Technology available for the automated subject indexing of visual material is developing swiftly. This article will describe a selection of systems found in the literature for the reproduction of high-quality pictures and for manual computerized indexing. For the indexing of nonprint collections to be successful, a visual element in an index is as important, if not more so, than the printed system that has been devised to access the pictures, whether they are photographs, slides, film frames, or entire films. Five conventional retrieval systems are discussed: the Universal Slide Classification System (U.S.); PRECIS, as used by the National Film Board of Canada and British Universities Film and Video Council; RVD Film Access System (the Netherlands); a thesaurus and computerized records of the Picture Division, Public Archives of Canada; and Iconclass (Holland). Videotex, as a possible option for the reproduction, storage, and display of images, is included. Alpha-mosaics, represented by Prestel (U.K.) and Teletel (France), and alpha-geometrics, represented by Telidon (Canada) and Cyclops (U.K.), are examined for the comparative quality of their graphics. Videodisc applications and advantages as well as four systems and projects are mentioned. Digital storage and one pilot project are also noted. Finally, four successful combinations of technologies are briefly presented.

At least two problems must be resolved before the indexing of visual materials can proceed smoothly and efficiently. First, because a subject authority for visual images, such as works of art, does not exist as it does for print works, the characteristic of "aboutness" can be much more difficult to determine for the visual media, especially for film segments and other components that form a unified whole. The second problem is the shortage of standardization by which subject indexing of images may be guided. Without the creation of universally accepted standards, networking is limited.

In consideration of the first problem, it was stated that conventional access tools and abstracts are inadequate because they depend on guidelines for printed items.[1]

Diane Brooks is with the Cataloging Department, Dalhousie University Library, Halifax, Nova Scotia.

The creation of appropriate abstracts for visual material involves "surrogates which will represent the differences in structure and extra-topical aspects of film and video works by which a user might make a relevance judgement."[2] O'Connor suggests that abstracts for films are ineffective due to lack of a photograph in the catalog entry. Photographs derived from frames that sufficiently represent the content of the film could be surrogates in the catalog that substitute for whole items. Until the mid-1970s, photographs were seldom included in indexes, with the exceptions of the 1927 illustrated catalog of the Fogg Art Museum and the 1946 pictorial index of the Library of Congress. The cost of developing and maintaining a visual index must be balanced by its contributions.[3]

The second problem, involving lack of standardization in visual indexing, has been expressed by the comment that

> in those few cases when subject control of a collection has been achieved the schemes utilized are usually idiosyncratic and private, preventing meaningful networking or sharing of iconographic information.[4]

Almost as many indexing systems exist as visual collections, and a survey of slide collections was conducted in 1968 by Irvine.[5] An anticipated result of this survey was the publication of a comparative study manual to enable slide librarians to select from a plethora of systems and procedures in existence. Irvine stated that

> the proposal of an "ideal" system is not the aim of the study, since most users of such a study will be working from already existing collections which have probably grown to such proportions that a complete revision would not be feasible.[6]

At the time of Irvine's 1971 article, the Bibliographic Systems Center at the School of Library Science, Case Western Reserve University, maintained lists of classification systems and subject headings in nearly every subject area. A guide to that collection, called *The Bibliography of Selected Material in Classification*, compiled by Barbara Denison, was then being published.[7] Irvine's survey indicated that 22 percent of U.S. respondents had adapted the classification systems of the Fogg Art Museum or of the Metropolitan Museum of Art created by Priscilla Farah, to their visual collections.

While the practicality and cost of converting large collections to a new indexing system are serious concerns, the value of standardization to efficient and rapid indexing and to the sharing of records is difficult to deny. This observation was supported by Gilley's 1973 survey in which respondents from Canadian and American museums verified the desire for greater standardization in classification and retrieval systems of slide collections.[8] At that time separate systems existed for art, history, and science slides.[9]

The requirements of an ultimate system for visual materials were suggested by Roberts, who stated that

> the ideal network must not only provide images of good quality but must also provide a verbal account that identifies and describes each work of art and allows for retrieval by indexed characteristics.[10]

This ideal network must hold a complete record of all pertinent information; must permit information to be upgraded and supplemented; must enable the retrieval of works of art in various locations or commissioned by the same patron; must include exhibition records as well as records of a particular artist's complete works or all works of comparable features; must represent the aesthetic, religious, social, political, and commercial functions of art works; and must permit subject or iconographic indexing. A visual database must be immense to be comprehensive if the estimate of 10 million works of art in the world is taken as conservative.[11]

On a smaller scale, concerning the creation of indexes, other points were suggested as important. Profiles of potential users should be prepared, and the necessary benefits of the system should be determined. It would also be helpful to determine the most efficient media for users and to acquire components accordingly, while anticipating the system's growth rate.[12] A literature search revealed that other factors were considered important for automated systems of nonprint collections and were mentioned time and again: flexibility for the addition of special search capabilities;

ease of information dissemination to inquirers; thoroughness and comprehension in providing title and subject access; ease of revision, use, and maintenance; a minimum of cost; and the ability to browse. Advantages of computerized indexing and retrieval were found to be faster coding of material; cross-tabulations between fields; possible studies of the frequency of appearance of certain characteristics; possible output in a variety of forms (terminal screen, microfiche, or abstracts) and in multiple copies; efficient searching, deletion, and correction of data; reductions in staff and cost; and rapid inventory of the collection. Disadvantages were perceived to be greater expense for equipment than manual systems, lack of appropriate hardware, and unpredictable downtime.

CONVENTIONAL RETRIEVAL SYSTEMS

Prior to late 1973 there appeared to be only one complete automated system for nonprint material in operation. This was the Universal Slide Classification System with Automatic Indexing, developed by Simons and Tansey at the State University of California, Santa Cruz, and published in 1969. It was the only one in a 1973 survey that could deal with art, history, and science slides.[13] The Universal Slide Classification System consists of individual subject divisions for art, history, and science with the first two divisions further broken down chronologically then geographically; the geographical divisions are designated by the numbers 002 to 999. Other subdivisions are alphabetically arranged. The art division, which covers both fine arts and architecture, includes period, place, medium, style, origin, and subject. At the time of LoPresti's 1973 article, only the art division was complete and was available from the Santa Cruz Master File, which held 40,000 slide records and an authority file of artists and architects on magnetic tape.[14] Nothing more recent about this system was located in the literature. The history division includes period, place, and subject. The science division indexes the natural elements. The system allows for fifteen sort fields, enabling the creation of a variety of indexes. Other advantages are simplicity, speed, and directness in indexing and retrieving slides, as well as the possibility of updating by the addition of new records and correction of old records.[15] By providing a format for slide indexing, the Universal Slide Classification System facilitates the sharing of records if included in a cooperative network.

In the mid- to late 1970s catalogers in Canada began seriously investigating PRECIS (PREserved Context Index System) for its precise, syntactical natural language treatment of subject indexing. Briefly, the procedure involves creating an abstract or subject statement for the material to be indexed, assigning PRECIS role operators and codes to the selected subject terms, and utilizing the computer to process the index records. Subject terms may also be enhanced by the addition of "see" references and "related terms" by using the British Library PRECIS thesaurus and the National Library of Canada and Library of Congress authority files. An advantage of PRECIS is that it provides hierarchical reference structure, preferred terms, and guidelines for their use without limiting itself to a predetermined vocabulary. Other advantages are its flexibility in dealing with any form of material; consistency and predictability for indexing and retrieval; and fast, efficient, and highly relevant subject searching. PRECIS is used by the National Film Board of Canada (NFB) for their FORMAT database of Canadian AV materials. That the indexing language is applicable in English and French is a perceived advantage of PRECIS. Records have about thirty information fields searchable on eighteen criteria by Boolean methods. Subject searches are achieved by three methods: the index string for precise retrieval; broad subject terms that permit browsing; or free-text scanning of the abstract for maximum comprehensiveness.[16] Abstracts are created for the document as a whole, rather than at the level of depth indexing for sequences or subthemes. This policy has been explained by promoting

> a reduction in indexing costs by defining a level of exhaustiveness in subject analysis that meets the needs of the index users while ensuring a high reuse rate of index strings.[17]

The PRECIS thesaurus is to be integrated online to enhance subject searching. A significant advantage to cost, speed, and efficiency is the fact that FORMAT's print catalogs are directly prepared from the database, rendering manual typesetting unnecessary. The MARC format is utilized to enter PRECIS subject authority records into bibliographic files, and the cooperation of this practice with the spirit of standardization is expressed by the statement that "a centralized online subject authority file is critical to achieving consistency and coordination in a decentralized indexing operation."[18] Perhaps the only deterrent to the wider use of PRECIS for indexing in North America is that its software is available only through UTLAS. A minor drawback may be the belief that training and use are difficult and that start-up and maintenance costs are high.

Another user of PRECIS is the HELPIS database, produced by the British Universities Film and Video Council. The file, input on the British Library's BLAISE-LINE, accesses AV materials in the U.K., and PRECIS provides a subject index for the printed and microfiched catalogs.[19] For additional information, see a review article by Mahapatra and Biswas on the major publications dealing with PRECIS up to 1982.[20] Dykstra published an introduction to the system for students and information professionals in 1985.[21]

In contrast to the NFB's approach to indexing, based on summarization of the entire document, is the RVD Film Access System designed by the Film Archive of the Netherlands Information Service.[22] In-depth film documentation had apparently not been developed elsewhere before this system was begun in 1978 and the input hardware was installed in 1981.[23] Documentation consists of recording all "film objects" and defining limits of a "scene," both of which are left to the discretion of specialists to define, with the goal of recording the maximum number of practical details. The contents, expressed at different levels of specificity, and the physical data for each film fragment are recorded on computer input forms. Specialists were selected who would be the most capable of identifying film content. The RVD documentalist verifies the specialists' work before the information is added to the database. The computer software used is the IBM Storage and Information Retrieval System (STAIRS) because of its flexibility, need for minimum maintenance, and accessibility both by a thesaurus and free text. IBM software also counts the frequency of descriptors and creates broad, narrow, and related terms. Several forms of output are obtainable: the complete documentation online or on microfiche, an abstract from the film archive catalog, or bibliographic references and annotations to material in the original files.[24] The primary disadvantage of the RVD Film Access System is the long documentation procedure coupled with the vast amount of material to be documented in the film archive. One hour of film requires eighteen hours of documentation by the specialist and five hours of verification by the documentalist.[25] The archive consists of about six million meters of film. Before the analysis even begins, the 35mm films are transferred to videocassettes and two copies are made, one for use in the archives and one to be viewed by the specialist. Each cassette contains film fragments that are individually numbered and timed. The broad range of subjects covered by films in the collection also hampers the creation of a workable thesaurus for access to the database. As a result, an individual thesaurus must be prepared for each subject before incorporation into a general work. The OECD Macrothesaurus was selected as a guideline for the development of an RVD thesaurus.

Another thesaurus was being developed by the picture division of the Public Archives of Canada at the time of a 1980–81 article by Ohlgren. This experimental, automated thesaurus of iconographic terms will facilitate controlled indexing in Canadian museums and promote online access to information on about 100,000 works of art in the picture division. At that time costume and architecture thesauri had been published, and thesauri for activities, artifacts, flora, fauna, insignia, landscapes, people, and transportation were in progress.[26] The picture division also decided in the early 1980s to prepare new computerized records directly from artworks in the collection. Standards of description were devised after investigating cataloging systems in Canada, the U.S., Great Britain, and France, and sixty-eight fields, divided into archival, artist, physical, and historical sections, were selected. A record consists of "each discrete image that can stand on its own," reflecting an emphasis on subject indexing in the spirit of Foskett's depth indexing and similar to the RVD Film Archive's practice.[27] A future project of the picture division is to transfer the collection to videodisc.

A proposed alternative to thesauri and lengthy film documentation for iconographic indexing is Iconclass, developed and published at the University of Leiden, Holland.[28] Begun about 1947 by H. van de Waal, an abbreviated version was published in 1968, and the entire work of eighteen volumes, consisting of the classification, bibliography, alphabetical index, and user's manual, was expected to be completed by 1984. (Verification of its completion was not found in the literature.) The philosophy of Iconclass was such that

> iconographic entities are arranged into a more or less coherent whole in which—sometimes lengthy—descriptions not of actual works of art but of subjects, themes and motifs in abstracto are represented by codes.[29]

While the system is long and massive, its advantage is that the classification has all been predetermined. The classifier merely has to select the appropriate notation that applies to the artwork in question, like using LC or DDC for printed material. A disadvantage, from the point of view of this article, is that it was not created with automation specifically in mind. A notation composed of letters and numbers was assigned to each term or clause describing an iconographical entity. For example, the notation 25 G 3 (LAUREL)(+ 225) represents "withered branch of a laurel tree."[30] The motivation behind Iconclass is to transform all conventional iconographic descriptions of Western art into codes. The classification system contains nine major groups—religion and magic; nature; human being, man in general; society, civilization, culture; abstract ideas and concepts; history; Bible; literature; and classical mythology and ancient history—each permitting subdivision as needed. Like DDC, each additional subdivision makes the classification more specific. The advantages of Iconclass are its superior capacity to retrieve information, retrieval of relevant iconographic references, and allowance of more than one subject to be assigned to a specific work of art. The computer-generated alphabetical index lists keywords followed by the traditional iconographic descriptions using or suggesting the keywords and their assigned notations. Cross-references direct the user from one keyword to another. The bibliography is said to be the most extensive survey of iconographic literature ever published.[31] At the time of Couprie's 1983 article, as many as 200,000 artwork reproductions had been indexed by Iconclass. Among other Iconclass projects, the most important then in progress was the Marburger Index, involving the indexing of 150,000 photographs of German and foreign art in West Germany. In this project, various levels of specificity are indexed.

Another project that expected to use Iconclass at the time of Sunderland's 1983 article was the computerized indexing of 1,300,000 microfiche photographs and reproductions of European art (from c.1200 to the present) held in the Courtauld Institute of Art, London.[32] The database will be searchable and sortable by fields, and the subject index is meant to be beneficial to historians of any subject, not only art. It is expected that a thesaurus of controlled vocabulary (i.e., Iconclass) would facilitate use of the database more than would searching by free text and that the Witt Library would have to produce their own subject authority list because of the range of material to be covered. The primary disadvantage of the anticipated database is that it would not include visual images. To compensate, copies of the microfiche catalog have been distributed worldwide and would be numerically concorded with the database.

VIDEOTEX

The next section of this article concerns videotex' two designs—alpha-mosaics and alpha-meometrics—and how well each reproduces visual images. In addition, their indexing facilities will be discussed. Videotex has been included in a paper about computerized indexing of visual materials because it is seen as a possible option for the formation of accurate surrogates or substitutes for items in nonprint collections. Videotex is a potential method of fulfilling the need to express the "aboutness" of an object visually, rather than by a written or notational description. That is, a photograph, slide, film fragment representative, and so on can be seen directly on a videotex screen by which a researcher can access a collection. The practical utilization of videotex for this purpose will greatly depend on how sophisticated their indexing systems can be made or on the possibility of supplementing their indexing facility with standard bibliographic search systems.

While videotex in Europe was based primarily on alpha-mosaic systems, Canada and Japan developed alpha-geometrics with much greater graphic fidelity. The former system uses a grid of twenty-four lines, each with up to forty alphanumeric characters. Rather than a character, each line may display a smaller array of about six square cells, which may be in two of seven available colors. A drawback of this system for graphics is that the 5,760 possible graphic elements are insufficient to create curved or diagonal lines. Strings of graphics characters, selected as the best representation of an object, are very limited in detail and lack flexibility in the way that an image is produced on the screen.[33] Much of the need for fine graphics was met by dynamically redefinable character set (DRCS), which does not greatly increase the cost of standard alpha-mosaic terminals. DRCS' only major problem is its color limitation permitting only two colors. A solution may lie in exchanging resolution for additional colors.[34] An application of DRCS is in charts, maps, and schematics. For access in a tree structure, DRCS could be attached to various levels of the tree; with keyword access, a subset of DRCS could be attached to each keyword. In the early 1980s plans existed to use DRCS to improve the British Prestel's graphics by adding an inexpensive page store to the terminal, permitting reprogramming for a variety of new symbols. At that time British Telecom had introduced Picture Prestel hardware, permitting display of a video quality image in a section of the screen.[35] Much less has been published about videotex in recent years. Another technology that makes the use of full-color photographs in videotex more realistic economically is the combination of picture coding techniques (PCT), originally developed for digital television, teleconferencing, and surveillance, with less costly semiconductor memory. The use of PCT in videotex decreases the number of bits required, rendering the implementation of low bandwidth circuits practical. The opinion was expressed that graphics techniques using costly manual composition should not be preferred over automated picture generation.[36]

The graphics encoding of alpha-geometrics, a medium resolution system as represented by Telidon, can be used equally well for teletext or videotex, in addition to use in personal computers and for enhancing databases. Alpha-geometrics rely largely on the receiving terminal to produce graphics. This terminal is a microcomputer with read-only memory (ROM), random-access memory (RAM), and a general purpose microprocessor. The desirability of a graphics-coding scheme accommodating terminal configurations of varying quality and costs while permitting maximum independence among terminals, databases, and communications media was discussed in the literature.[37] In alpha-geometrics, coordinate specifications—or picture description instructions (PDI)—direct the display process by describing five fundamental geometric shapes from which all images are created. Possible resolution is about 250 horizontal by 200 vertical separately resolvable points, which are each capable of assuming any of eight colors and six shades of gray. Display of quasiphotographic images, resembling facsimile, is also possible, and PDIs can be automatically given from a television camera's video output. Advantages of Telidon graphics are the following: very high image fidelity; possible image overlays and basic animation; various ways of producing images; storage of images of any digital medium, including magnetic disc, audio magnetic recording tape, semiconductor memory, and videodisc; transmission of images including detailed color pictures over any medium, such as wire pairs, cables, optical fibers, microwave links, line-of-sight laser links, and satellites, with transmission speed not inherently restricted; display of images on a selection of electronic media without altering the data to suit the medium; and a design to avoid obsolescence.[38] In fact, due to the design, Telidon graphics would be improved by broader-band transmission systems, higher-resolution displays, and more sophisticated computer memories. The Telidon philosophy was described as "extensibility of features while maintaining compatibility with established databases and user hardware."[39] The Bell System made several additions to Telidon's graphics encoding, including alpha-mosaics that permit access to data encoded in the European format, macro-PDIs that can be used to display repetitive graphics, a new PDI known as the increment (which allows definition of freehand curves such as signatures), and more colors.[40]

Alpha-geometrics were adopted for the design of second generation videotex in Europe. The British alpha-geometric color videotex system, Cyclops, developed by the Open University to operate on a BBC micro, is a close match in quality with Telidon.[41] Cyclops is an eight-color, 240-by-320-point graphics terminal compatible with Prestel.[42] The Open University developed their own viewdata system, called Optel, on their DEC-20 computers rather than use Prestel because of the latter system's limitations in rapid access and access paths. Like Prestel, the smallest unit of random-access retrieval in Optel is the page, each of which has a name and is divided into any number of frames that are identified numerically. Keywords are assigned to each page; that is, accessing a keyword retrieves a list of pages citing that keyword. A disadvantage of Optel is that bibliographic retrieval has not been incorporated in order to meet their design guideline of expending the minimum CPU time on retrieval requests. In spite of this limitation, Optel's command interface is an amalgam between videotex (Prestel) and a CPU time-sharing system (a Univac mainframe).[43] The Optel database contains primarily text. The facility for visual material was unproven at the time of Bacsich's 1981 article, as was the keyword retrieval. Due to a scarcity of current published articles about videotex, the quality of these facilities is still unknown to the author.

While Telidon had adopted a computer graphics system as its model, rather than the teletype machine that Prestel and Teletel (France) were founded on, the indexing was still classificatory like the alpha-mosaic systems. In order to accommodate new information in the database, the classification had to be expanded and reorganized. Also, since information was stored in arbitrary precoordinate indexes that may not coincide with the hierarchy of numbered pages, use of Telidon may be hampered by a very large database. Telidon and Prestel use numeric keypads for information access. All three systems, Telidon, Prestel, and Teletel, would have to be redesigned in order to achieve postcoordinate keyword searching like online systems such as DIALOG.[44] Evidence in the literature suggests that redesign is taking place, for example, a 1981 publication about the Montreal keyword index for Telidon and a keyword index for Teletel.[45] Teletel utilizes alpha-numeric search strategies and permits keyword access, as opposed to hierarchical retrieval structures featured by the other two systems. Indexing featured in Teletel's STAR nodes, which was experimental in the early 1980s, was more sophisticated due to the organization of its database and its keyword dictionaries that permitted flexible searching. It was said about Teletal that

> although the data structuring facilities available in videotex systems are still relatively primitive and need more attention, the Teletel/Star experimental system has shown that simple concepts—the use of keywords and the association of an executable program with each page—can lead to far more sophisticated, user-oriented facilities.[46]

While STAR permitted many indexes, they continued to be precoordinate indexes, precluding retrieval of synthesized information that is not predetermined. Furthermore, it is unknown to this author what quality of graphics has been achieved by Teletel.

In order to achieve multiple hits with keyword searches or postcoordinate searching, all three videotex systems would have to undergo redesign. The computer technology necessary for such a relational approach existed in the late 1970s.[47] The opinion was expressed that the first generation of videotex systems will be limited to trivial information retrieval or to repetitive access to specific, frequently updated information.[48] The value of an interface between videotex and more sophisticated bibliographic search systems was endorsed by Tompa, Gecsei, and Bochmann who stated that

> the experience gained by those who have dealt with conventional programming languages, graphics, operating systems, and database systems should be applied to the future design of videotex systems. . . . In addition, the marriage of page-oriented videotex technology with conventional record-oriented retrieval systems seems very promising.[49]

VIDEODISC AND DIGITAL STORAGE

The videodisc offers advantages to visual indexing due to its high-density storage and rapid random access. Europe and the U.S. pioneered development of the videodisc, with Japan contributing research later. In the late 1970s systems allowed only playback, but work was being done to achieve recording. Recording on and playback from discs can be accomplished by four major techniques—mechanical, capacitive, optical, and magnetic. System examples of the first three categories are TeD (formerly Teldec), a system developed jointly by AEG-Telefunken of West Germany and Decca of the U.K. (mechanical); JVC (or the Victor Co. of Japan) and RCA of the U.S. (capacitive); and Philips/MCA and Thomson-CSF, a French manufacturer (optical).[50] In the early 1980s there were no magnetic systems ready for commercial distribution. However, magnetic systems were being used for broadcasting of freeze-frame and slow-motion playback in industry and medicine. While the videocassette recorder has the recording advantage, discs are advantageous in that videodisc players and the discs themselves are less costly than the VCR. Also, discs are less likely to be copied than cassettes, rendering them more secure. An application of videodiscs is as an automatic storage and retrieval system, likely in conjunction with minicomputers or mainframes. IBM and Xerox are developing the videodisc. Becker provided a concise, useful outline of definitions, questions, and answers concerning optical, digital, and compact discs.[51]

A particular optical videodisc system, the Discovision Industrial Player developed by MCA for educational and instructional applications, was described as "a microprocessor-based, programmable audio-visual information system."[52] An optical videodisc player equipped with a low-power laser relays any combination of moving images, still frames, and sound from a prerecorded videodisc to any video projector, video monitor, or television receiver. A disc's storage capacity is a maximum of 54,000 frames of information on each side, producing 30 minutes of playback information. The principal feature that makes optical videodiscs important for storage and retrieval is the numbering of each frame stored on a disc, allowing automatic accessibility by frame number via a remote control unit. Frame accessing takes 2.5 seconds on an average, or 5 seconds maximum when searching from end to end. The Discovision system uses a specialized type of analog recording that boosts storage capacity of one side of a 30cm disc to 60 billion bits of information.

Two techniques for high-resolution document storage are matrixing and scan conversion. High resolution may be defined in terms of the number of pixels per television frame. The conventional NTSC (National Television Systems Committee) TV frame, used for U.S./Canadian broadcast television, can display 2.0 by 10^5 pixels per frame. Documents requiring greater than this number of pixels will fill more than one television frame, necessitating the use of suitable interface electronics. For example, an $8^1/_2$-by-11-inch typewritten page needs about 1.6 by 10^6 pixels, or eight TV frames, while a 2,000 line photograph needs 4 by 10^6 pixels, or about 20 TV frames.[53] In late 1985 the high definition television (HDTV) videodisc was introduced. With its superior resolution of 1,125 lines, this innovation became more suitable for accessing visual materials.[54]

Matrixing is the most basic method of multiframe recording for documents exceeding 2.0 by 10^5 pixels. A document is divided into a matrix of postcard-size subpictures, each of which can be individually scanned and recorded as one television frame. To facilitate indexing during playback, various plans or overviews of the computer document and its quadrants may be recorded on single television frames. A disadvantage of the matrix technique is the formation of artifacts at matrix boundaries, caused by photographically reconstructing the document from transparencies of each subpicture. The second method of high-resolution document storage, scan conversion, almost eliminates boundary artifacts while permitting a full resolution display.[55]

An experimental optical disc program ran from October 1, 1982, to December 31, 1985, at the Library of Congress.[56] As the first genuine operating system, the nonprint section of the experiment stored and accessed motion pictures and stills using analog-signal laser videodiscs and computers. Six discs, containing material from the prints and photographs division and the motion picture, broadcasting, and recorded sound division, were produced for visual images, amounting to more than 150,000 images. One part of the program created new records, called captions, structured according to a preliminary version of MARC for visual image collections, and tested the automated retrieval of nonprint material. The picture captions—detailed lists of information about an image, including subject terms—formed the searchable database. To locate a picture featuring a specific subject, the search was undertaken by the computer, which then directed the videodisc player to show the pictures. Or, the pictures could first be browsed, and the computer directed to show the desired captions. The software for the search and retrieval system already existed commercially, making its expensive and time-consuming in-house development unnecessary. The Library of Congress plans to build the database gradually as uncataloged material is processed, to produce more discs if the experiment is judged a success, and to share the database and discs. The primary advantage of videodiscs is vastly improved access.[57] The project continued to operate in 1986, and more information about it may be obtained from McQueen and Boss' library technology update.[58] This report also discusses in detail videodiscs, optical digital discs, CD-ROMS, and their library applications.

The National Library of Canada (NLC) created a videodisc system for promotion and instruction. One side of the disc stores a film about the NLC and slides about the Canadian national bibliography. The other side contains a narrated tour of the library. A Discovision videodisc player and an Apple II microcomputer were combined to achieve the library's goal of "a fully interactive disc system in order to demonstrate the information storage and retrieval capabilities of combined videodisc/microcomputer systems."[59] The project managers concluded that the analog nature of optical videodiscs is better suited to audiovisual storage than to text. Also, because of the expense and lack of updating capability, the videodisc is best for multipurpose visual databases.[60]

Another project using videodiscs and completed in December 1985 was the U.K. Domesday Survey, supervised by Howard Newby at the University of Essex.[61] The survey was undertaken on behalf of the BBC for a thorough statistical profile of Britain in the 1980s on the occasion of the Domesday Book's 900th anniversary. Information for the videodisc database was supplied in map form, and the discs contain indexes for retrieving and presenting information. The two videodiscs were published in November 1986 and are said to be the first integrated database for public access.[62] They combine information in graphical and textual format from a large number of subject sources. Retrieval is accomplished by keyword and hierarchical indexes, which were described by Lee.[63]

Laser optical storage and retrieval became more accessible to smaller libraries when Bibliofile was introduced on CD-ROM by The Library Corporation in 1984. Using an IBM PC this system accesses about 500,000 MARC catalog records. CD-ROM holds promise for storing pictures since 600mb, representing one billion characters—about 200,000 pages of text or about 1,000 images—will fit on one disc.[64] To overcome CD-ROM's drawback of difficult access to the vast quantity of information stored on it, McConnell suggested a revolutionary indexing scheme known as hypertext. Unlike conventional retrieval systems that may lead a searcher to a dead end, a hypertext link to another location in the database is always available. McConnell affirmed that "hypertext can support a very personal intellectual journey, with all items related in a way that is both logical and intuitive, the way the human mind thinks."[65]

Digital storage technology is likely to be important in the future for computerized visual indexing and retrieval. Optical disc storage may render information more difficult to retrieve on the spot than digital storage, making the latter preferable for academic purposes. It may also be more secure, stable, and have a greater degree of permanence for art history purposes. Electronic images can be stored on computer and transmitted by satellite or fiber-optic cable anywhere in the world. It will also be possible to produce electronic digital images with greater color fidelity and resolution than can be produced by conventional color photography. This technique of image dissemination will be enhanced by the variety of methods for image retrieval and display, such as a screen, desk monitor, or color prints and slides. In art galleries or libraries, images of artworks will be available immediately. The primary drawback of electronic imaging technology is the high cost, which may be prohibitive to any single academic profession. To be practical economically, the system may have to be universal in its coverage, so that not only art, history, and the sciences but also medicine, publishing, and manufacturing would use one standard system, thus obeying the principle of standardization on a grand scale.

Fontaine described the problems involved with digitizing images by microcomputer through a time and motion study. Photographs were digitized with a Datacopy Corporation Integrated Imaging System and an IBM PC/XT at the University of Texas. The study did not find digitizing more efficient than microfilming, but the author pointed out that the process was slowed by the limitations of their hardware and software. In contrast, the Grinnell Image Processing system and a DEC MicroVax can digitize 60 frames per second. Storage involves 260kb for black-and-white images or 788kb for color, and images are stored on magnetic disc.[66]

An effort is being made at Birkbeck College (University of London) to launch an electronic digital imaging project.[67] The first phase will study the recording of images, with storage and retrieval to follow. One aim of the pilot project is the development of indexing systems for speed, ease, and efficiency of access that will match the quality of the image itself. A major obstacle to the program is the massive amount of information to be stored.

SYSTEM-SYSTEM INTERACTION

This final section will briefly describe four combinations of digital recording, videodisc and videotex storage, computer memory, or television display for storing and accessing moving and still images. There are likely many more examples in the literature, but those presented should be sufficient to stimulate consideration of other possible interactions between and among technologies to promote improved visual representation.

The importance of digital recording techniques to the use of videodiscs as computer mass memories was discussed in the literature.[68] Battelle-Northwest designed a videodisc system that implements digital recording. Digital Recording Corporation, the system's sponsors at the time of Horder's 1979 report, claimed that when photographic material was used as the recording medium, digital recording achieved an information density about seven to eight thousand times greater than 20X microfilm.[69] The Fugi Photo Film Company was investigating the system as a replacement for microfilm, and its application to mass data storage was being looked into by Bell and Howell's datatape division.

The National Agricultural Library (NAL) in Beltsville, Maryland, combined digital text and analog graphics into one database, while maintaining rapid retrieval. The library uses an IBM PC/XT and a Pioneer LDV-1000 videodisc player connected by a LaserData controller with interface card. This system is used in combination with BRS/Search, which provides retrieval capabilities. LaserData's modifications enabled BRS/Search to access graphics while indexing and retrieving information from the full-text database. Subject retrieval is by menu, any word in the text can be accessed, and Boolean operators can be used.[70]

Dataland is an experimental information system developed at MIT in the late 1970s.[71] The prototype consisted of a wall-sized, full-color digital television graphics "window" with synchronized stereo sound to create a spatial environment. The television's "visual repertoire" was supported by an optical videodisc as a storage medium, thus lowering costs. Dataland organized and accessed data according to the spatial sense of the user who may travel over this environment by means of joystick control, in the same sense as locating items manually on a desk. Various forms of visual data, including television, films, slides, photographs, and maps, could be located through the window, zoomed in on, and interacted with. Information forming the images was stored in the computer's "frame buffer." Anything that may be put on television—live, recorded, closed circuit, or broadcast—could be put in Dataland's graphics window. The system's potential is that

> the ensemble of data types and associated interactions enumerated here form an initial set and by no means exhaust the long term potentialities of user-system interaction.[72]

The combination of material recorded on videodisc and videotex and displayed on one color television monitor was reported by Williams.[73] This interactive system was developed by British Leyland's software subsidiary, Istel Ltd. Moving or still images may be shown together with text and sound, data may be updated, and the system requires neither keyboard nor computer skills to be used. Equipment used is a European standard (PAL) Philips Industrial Disc Player with an RS232 interface permitting use with a BBC Acorn micro. While the system is used for commercial and educational purposes, it could have potential applications in art collections.

Chen published an overview of optical videodisc technology and how it may be combined with microcomputers.[74] In this article specific projects were discussed, such as the National Library of Medicine's program to store digitized images on optical videodiscs controlled by computer. Chen also listed commercial videodisc and CD-ROM systems.

In 1987 General Electric Company announced an innovative technique of combining "the interactivity of computer graphics and text retrieval with the realism of television video and digital stereo sound."[75] The system, which depends on compressing data by a factor of 120:1, is used with an IBM PC/AT. The company anticipates that development projects will use their media independent digital video interactive (DVI) technology in early 1988.

CONCLUSION

In indexing visual materials, surrogate images substitute for the physical pictures, while an indexing language pulls together the surrogates on a particular subject. This duality of purpose has been expressed by Roberts:

> New film, improved cameras, analog and digital computers and laser beam technology all allow for a greatly improved imagery and distribution of imagery. The development of thesauri, indexing languages and improved computer programming allow for better intellectual control of databases.[76]

Both the production of good quality images and an automated system that provides accurate subject retrieval are important.

Some of the best conventional retrieval systems and the few complete optical disc systems operating in libraries were described. A selection of system interfaces were offered to illustrate how combinations have been achieved. Videotex was discussed as a possible option for storing and reproducing pictures. It was noted that alpha-geometrics can be used to enhance databases and can be transmitted over any medium, permitting remote access. Videodiscs can be used to store alphageometric images, and computers can store electronic digital images. Minicomputers or mainframes have been used in tandem with videodiscs as automatic storage and retrieval systems. Videotex and CPU timesharing have been merged. The combinations seem endless. Videotex reproduces images with varying degrees of success, but their indexing systems are limited. They need to be interfaced with bibliographic, postcoordinate keyword search and retrieval techniques that have been developed for print collections. There is some evidence that this process has begun.

ACKNOWLEDGEMENTS

The author wishes to acknowledge Professor Jack Meadows, currently of the Library and Information Studies Department, Loughborough University, England, who initiated the research contained in this paper. In 1986 the author undertook a literature search and telephone interviews on the topic of computerized indexing, storage, and retrieval of graphics on behalf of the Primary Communications Research Centre, University of Leicester, England, for an ongoing project funded by the British Library.

REFERENCES AND NOTES

1. Brian C. O'Connor, "Access to Moving Image Documents: Background Concepts and Proposals for Surrogates for Film and Video Works," *Journal of Documentation* 41:209–20 (Dec. 1985).
2. Ibid., p.209.
3. Elizabeth Matthew Lewis, "Visual Indexing of Graphic Material," *Special Libraries* 67:518–27 (Nov. 1976).
4. Thomas H. Ohlgren, "Conference on Computerized Inventory Standards for Works of Art: A Review," *Visual Resources* 1:200–203 (Fall 1980/Winter 1981).
5. Betty Jo Irvine, "Slide Classification: A Historical Survey," *College & Research Libraries* 32:23–30 (Jan. 1971).
6. Ibid., p.24.
7. Ibid., p.25.
8. Beulah L. Gilley, "Declassifying the Slide Secrets," *Museum News* 52:45–48 (Mar. 1974).
9. Ibid., p.47.
10. Helene E. Roberts, "Visual Resources: Proposals for an Ideal Network," *Art Libraries Journal* 10:32–41 (Autumn 1985).
11. Ibid., p.40.
12. Lewis, "Visual Indexing," p.525.
13. Gilley, "Declassifying," p.47.
14. Maryellen LoPresti, "An Automated Slide Classification System at Georgia Tech," *Special Libraries* 64:509–13 (Nov. 1973).
15. Ibid., p.512.
16. Donald Bidd, Louise de Chevigny, and Margo Marshall, "PRECIS for Subject Access in a National Audiovisual Information System," *Canadian Library Journal* 43:177–84 (June 1986).
17. Ibid., p.180.
18. Ibid.
19. Elizabeth Oliver, "Audio-visual Materials and the Law Librarian," *Law Librarian* 16:81–83 (Aug. 1985).
20. M. Mahapatra and S. C. Biswas, "PRECIS: Its Theory and Application—An Extended State-of-the-Art Review from the Beginning up to 1982," *Libri* 33:316–30 (Dec. 1983).
21. Mary Dykstra, *PRECIS: A Primer* (London: British Library, Bibliographic Services Division, 1985).
22. R. H. J. Egeter Van Kuyk, "Historical Film Documentation at the Netherlands Information Service," *UNESCO Journal of Information Science Librarianship and Archives Administration* 3:227–34 (Oct.-Dec. 1981).
23. Ibid., p.229.
24. Ibid., p.234.
25. Ibid., p.230.
26. Ohlgren, "Conference on Computerized Inventory," p.202.
27. Ibid.
28. L. D. Couprie, "Iconclass: An Iconographic Classification System," *Art Libraries Journal* 8:32–49 (Summer 1983).
29. Ibid., p.33.
30. Ibid., p.36.
31. Ibid., p.45.
32. John Sunderland, "The Witt Library Getty Project," *Art Libraries Journal* 8:27–31 (Summer 1983).
33. Ian McLaren, "Graphics and Presentation: If This Is the Shape of Things to Come, Why Does It Look Nasty?" *Aslib Proceedings* 35:319–24 (Aug. 1983).
34. O. Lambert and others, "Antiope and

D.R.C.S.," *IEEE Transactions on Consumer Electronics* 26:600–604 (Aug. 1980).
35. Kevin Smith, "Prestel to Transmit Colour Photographs," *Electronics* 53:78, 80 (Oct. 23, 1980).
36. K. E. Clarke, "The Application of Picture Coding Techniques to Viewdata," *IEEE Transactions on Consumer Electronics* 26:568–77 (Aug. 1980).
37. H. G. Bown and others, "Comparative Terminal Realizations with Alpha-geometric Coding," *IEEE Transactions on Consumer Electronics* 26:605–17 (Aug. 1980).
38. Andrej Tenne-Sens, "Telidon Graphics and Library Applications," *Information Technology and Libraries* 1:98–110 (June 1982).
39. Ibid., p.107.
40. Ibid., p.108.
41. McLaren, "Graphics and Presentation," p.320.
42. P. Bacsich, "The Open University Viewdata System," *Viewdata: Conference Proceedings, Oct. 1981* (Pinner, Middx, Eng.: Online Publ., 1981), p.647–57.
43. Ibid., p.652.
44. John Kuo-chang Wang, "Telidon: The Canadian Videotex System," *Journal of Educational Media Science* 19:336–43 (Summer 1982).
45. F. W. Tompa, J. Gecsei, and G. V. Bochmann, "Data Structuring Facilities for Interactive Videotex Systems," *Computer* (U.S.A.) 14:72–81 (Aug. 1981).
46. Ibid., p.80.
47. Diane C. P. Smith and John Miles Smith, "Relational Database Machines," *IEEE Computer* 12:28–38 (Mar. 1979).
48. A. J. S. Ball, "Videotex: Chimera or Dream Machine," *Canadian Library Journal* 38:11–15 (Feb. 1981).
49. Tompa, Gecsei, and Bochmann, "Data Structuring," p.80.
50. Alan Horder, "Videodiscs—Their Potential for Information Storage and Retrieval," *Reprographics Quarterly* 12:46–48 (Spring 1979).
51. Karen A. Becker, "CD-ROM: A Primer," *College & Research Libraries News* 48:388–93 (July/Aug. 1987).
52. Horder, "Videodiscs," p.47.
53. George Kenney, "Special Purpose Applications of the Optical Videodisc System," *IEEE Transactions on Consumer Electronics* 22:327–38 (Nov. 1976).
54. William R. Nugent, "Optical Discs—An Emerging Technology for Libraries," *IFLA Journal* 12:175–81 (Aug. 1986).
55. Kenney, "Special Purpose Applications," p.334.
56. Carl Fleischhauer, "Optical Disc Pilot Program," *Library of Congress Information Bulletin* 42:312–16 (Sept. 12, 1983).
57. Ibid., p.314.
58. Judy McQueen and Richard W. Boss, *Videodisc and Optical Digital Disk Technologies and Their Applications in Libraries, 1986 Update* (Chicago: American Library Assn., 1986).
59. A. M. Hendley, *Optical Disk Systems, A Guide to the Technology and the Applications in the Fields of Publishing and Library and Information Work* (The Hague, Netherlands: International Federation of Library Associations and Institutions, IFLA Professional Reports, No. 12, 1986, ERIC ED 278 399).
60. Ibid., p.35.
61. *Current Research in Library and Information Science* (Sept. 1985), abstract no.53.
62. Richard Tapper, "Building the Domesday Database—Lessons for Integrated Database Development," *Aslib Proceedings* 39:107–21 (Apr. 1987).
63. David Lee, "Indexing the Domesday Project," *The Indexer* 15:145–50 (Apr. 1987).
64. Ching-chih Chen, "Libraries in the Information Age: Where Are the Microcomputer and Laser Optical Disc Technologies Taking Us?" *Microcomputers for Information Management* 3:253–65 (Dec. 1986).
65. Karen McConnell, "Optical Technology: Interacting with Traditional Systems," *Wilson Library Bulletin* 61:21–24 (June 1987).
66. Claudia Fontaine, "Time Study of the Digitization of Images in a Microcomputer Environment," *Microcomputers for Information Management* 4:129–38 (June 1987).
67. Unpublished material and discussion with chair of the advisory group at Birkbeck College, University of London.
68. Horder, "Videodiscs," p.46.
69. Ibid., p.48.
70. Pamela Q. J. Andre, "Full-Text Access and Laser Videodiscs: The National Agricultural Library System," *Library Hi Tech* 4:13–27 (Spring 1986).
71. Richard A. Bolt, *Spatial Data Management Systems* (Cambridge, Mass.: MIT, Architecture Machine Group, 1978, NTIS report AD-A070 243).
72. Ibid., p.49.
73. Bernard Williams, "Combined Videodisc/Videotex at British Leyland," *Reprographics Quarterly* 17:68–69 (Spring 1984).
74. Ching-chih Chen, "Micro-based Optical Videodisc Applications," *Microcomputers for Information Management* 2:217–39 (Dec. 1985).
75. Tom Manuel, "GE Takes the Wraps Off Its Interactive Video System," *Electronics* 60:97–99 (Nov. 26, 1987).
76. Roberts, "Visual Resources," p.33.

Serials Conversion at the University of Massachusetts at Amherst

Patricia Banach and Cynthia Spell

From November 1984 through December 1985, OCLC retrospectively converted more than 27,000 serial titles held by the University of Massachusetts Library at Amherst. Details of the reasons for choosing OCLC, the contract, and results are summarized. Although final conclusions as to the success of the project are still being formulated, the library's goal of a timely and high quality conversion of a large number of serial records seems to have been achieved.

BACKGROUND TO THE CONVERSION PROJECT

The University of Massachusetts Library at Amherst serves a total student population of 26,450, of whom 6,700 are graduate students. The library system consists of a main library and three major branch libraries: biological sciences, physical sciences, and music. The collection totals over 2 million bound volumes and 1½ million microforms. All technical processing is done centrally at the main library.

The total serials collection today stands at approximately 16,000 currently received titles and 21,600 noncurrent, for a total of 37,600. In spring and summer 1984, the period in which we were faced with our serials recon decision, the serials collection was somewhat less: 13,500 currently received serials and 21,000 noncurrent, totaling 34,500. (These figures include almost 3,000 monographic series and analyzed serials not part of our serials recon project, since they are considered monographs.)

All of the serials and periodicals in the collection are fully cataloged and classified according to LC. Even current periodicals are not shelved by title, except for popular titles available for browsing in a separate area of the current periodicals room. The library started using successive entry cataloging for serials title changes in spring 1971, shortly after the LC announced that it would implement this rule change mandated by *AACR1* in 1967. Titles cataloged prior to that date, which have not required recataloging for some other reason, remain cataloged according to latest entry principles.

From 1971 to February 1980, the library had its own in-house automated catalog card production system, using the MARC tapes and the university's computer. In February 1980 we began gradually switching to OCLC for cataloging of monographs, but it was not until July 1, 1984, that we first cataloged serials on OCLC. Consequently, although we had more than 400,000 cataloging records in machine-readable form, only a small fraction of monographic records and no serials at all

Patricia Banach is Head, Serials Automation Section, and Cynthia Spell is Serials Cataloger, Cataloging Department, University Library, University of Massachusetts, Amherst.

were on OCLC in spring 1984, when we began to look at future automation plans. Many serial records pre-dated our own automated system and were not in machine-readable cataloging form at all, and the machine-readable records in the local card production system were not in MARC format and could not be easily converted to MARC format.

The primary impetus for undertaking a large-scale serials recon project in spring 1984 was the library's participation with Five-Colleges, Inc., in acquiring an OCLC LS/2000 integrated library automation system. To load data into this system, it was necessary to undertake conversion.

The first phase of conversion consisted of contracting with OCLC to machine match our existing machine-readable cataloging records against the OCLC database and create archive tapes from these records. (Because of the structuring of our in-house cataloging system, our records were not in a form that could be loaded directly into LS/2000.) Only machine-readable records with Library of Congress card numbers were candidates for conversion during this process. For serials, that constituted only a very small percentage of the total serial collection. (These records then were skipped in subsequent conversion projects.)

The next phase of conversion was focused on records for our physical sciences branch library, the university library's test site for the LS/2000 system. In spring 1984 a project was being negotiated with OCLC to complete conversion of the physical sciences and biological sciences libraries' monographic collections, most of which had already been converted in-house.

As part of the on-going planning process for implementing cataloging via OCLC, an eight-member committee of technical services librarians was formed specifically to study the implications for serials. This committee undertook to formulate procedures for doing an in-house retrospective conversion of physical sciences library serial titles. Procedures were in fact drafted, and tentative plans made to assign a high-level paraprofessional to the project. But before actually beginning in-house conversion, the senior serials cataloger took a random batch of sixty-nine physical sciences serial records to the OCLC terminal and attempted to try out the devised procedures.

The serials cataloger determined that forty-eight of the sixty-nine titles had problems of one kind or another. Not all were serious, but after reviewing them, the committee that had drafted the recon procedure felt it would be very difficult to draft procedures to deal with every type of problem. Therefore, most of the records would have to be screened by a serials cataloger.

At that time our staff available to work on serials cataloging consisted of one senior serials cataloger; one monographic cataloger (relatively new to the staff), who was being trained to catalog serials; and one serials cataloging assistant, who can best be described as a high-level paraprofessional.

It became apparent that the serials recon project for the physical sciences library would be a very severe draw on the time of these serials catalogers, who still had to keep up with current cataloging and recataloging of title changes, etc. It was also apparent that the library was not likely to hire, train, and supervise additional staff for a serials recon project without again placing a significant work load on the existing serials cataloging staff.

At this point the committee considered three options:

1. To go ahead with the in-house conversion, necessitating either a large influx of staff or a very long conversion project that would not meet the goal of having the serials loaded to LS/2000 by the fall 1984 semester. (Funding for additional staff was problematic anyway, so this was not a particularly viable option.)

2. To do a quick and dirty conversion, which would make any subsequent linking project difficult and might take years to clean up.

3. To include serials in the monographic recon project already being negotiated with OCLC.

The committee concluded that the last option, that of contracting out the project to OCLC, was the way to go. It was felt that this would result in a larger number of records being converted by OCLC's already highly trained and experienced staff in a timely fashion. The proposed editing criteria for this project would require that all our cataloging points of access be present or added to the OCLC record. Consequently the fact that the subsequent LS/2000 record might not exactly match our card catalog, serial Kardex, or serials union list would be less problematic. Those records that OCLC returned to us as problems could be reconciled by the available serials cataloging staff, who would not be burdened with involvement in an in-house conversion.

CONTRACT WITH OCLC

The director of libraries agreed with the recommendation of the committee to contract out the physical sciences serials and the biological sciences serials, and thus there remained the matter of how to go about it. To insure that the records resulting from the OCLC recon project would be of high quality, a contract was negotiated with OCLC to define what constituted a valid matching record and what enhancements to records would be required. The technical specifications portion of the contract runs to eight pages, which is too lengthy to quote. Consequently, only the most important features are summarized below.

First of all, the contract did not call for production of new catalog cards. The library was seeking machine-readable records to be loaded into its online LS/2000 system. It was recognized from the outset that these records would not be identical to our hard-copy records but would represent the same bibliographic items.

A match of the following elements in our record and OCLC's was required:

1. Type of record
2. Bib level "S"
3. Successive entry required (S/L ent = "0")
4. Our library's title matching *any* of the following:
    a. Key title (222 field)
    b. Title proper (245 field)
    c. Varying title (212 or 246)
    d. Uniform title (130 or 240)
5. If any of the above titles matched, then OCLC was required to compare the additional bibliographic data (volume numbering, dates of publication, etc.).

Based on the above matching criteria, OCLC would establish a match where the bibliographic data reasonably supported that the two records being compared represented the same bibliographic entity.

If more than one OCLC record appeared to constitute a match, then the contract defined a hierarchy of records from which to choose based on the 042 field. That hierarchy was in this order:

1. Library of Congress or National Library of Canada authenicated record
2. other authenticated record as indicated by 042 field (e.g., nsdp, etc.)
3. unauthenticated record, the most complete and accurate as determined by OCLC staff.

Further, if our record was cataloged according to latest entry, OCLC was required to split it into as many successive entries as our holdings indicated. When no matching record existed in the OCLC online union catalog, the OCLC conversion staff was required to enter a new record based on our catalog card, changing our name headings to *AACR2* form if needed.

The contract also called for OCLC to enter two in-house ID numbers into separate 035 fields in the record. One ID number was that pertaining to our serials database. The other was the ID number from our automated catalog card production system.

Call number and holdings symbol were added to the online record, and OCLC was required to pencil the OCLC control number (or numbers, in the case of a latest versus successive entry split) on the shelflist card to be returned to us. We later added this number to our serials database.

Other additions to the OCLC online record required by the contract were as follows:

1. Our main entry had to be added to the OCLC record as an added entry if it was not already present in the record.

2. Any main or added entry in our record that did not match the *form* in the online record was to be checked in pencil and flagged as a name conflict. However, the record was to be converted with the form of entry found in the online record, provided that it was an LC record or CONSER authenticated, since this was presumed in the majority of cases to be the correct *AACR2*

form. If the online record was from another source (i.e., non-LC) and the entry had not been machine converted during the *AACR2* flip in 1980, then OCLC was required to change the entry in the online record to the form on our card.

Notes and subject headings from our records were also to be added by OCLC to the online record. Dates were edited in both fixed and variable fields. (The contract goes into additional detail, but these are the highlights.) Having negotiated a contract, the next step was to devise a plan to prepare the records to be converted for shipment to OCLC.

The library does not maintain a separate serials shelflist so the records to be converted had to be pulled from the shelflist. The committee charged with overseeing the conversion project had some concerns about sending to OCLC our main shelflist cards for physical and biological sciences serials and periodicals. It discussed what backup, if any, would be maintained. Although OCLC assured us that they had successfully converted millions of records without loss of any shelflist cards, we were of course apprehensive that their luck might change! The committee considered making photocopies of the shelflist cards but rejected that idea in favor of using brief computer-generated slips from our serial database.

These slips contained author/title data, call number, location, LC card number, ISSN (when available), our serials database ID number, and our holdings for the designated title. The holdings data were very important since many of our cataloging records contained only a general note concerning holdings. (The complete holdings are maintained on the serials Kardex.) Because the extent of holdings was necessary for OCLC to determine what and how many successive entries to create for our latest entry records, we produced the three-by-five slips in duplicate. One went into the shelflist as a replacement for the shelflist card, and one was paper-clipped to the shelflist card and sent to OCLC.

After the shelflist cards were pulled, and just prior to sending them to OCLC, the cards were reviewed by appropriate staff, and marginal titles were weeded, thus avoiding the expense of converting and maintaining material no longer deemed valuable to the collection.

All of this preliminary work was accomplished in late spring and summer 1984, and the shelflist cards were sent to OCLC in October of that year. OCLC began the conversion project on November 15, 1984, and completed conversion of 6,182 serial records by January 30, 1985. An additional 234 records were designated problems by OCLC and not converted, and 152 records were not touched, since OCLC found that they had been converted in-house during the course of the project. (These 152 were records that needed to be cataloged or recataloged during the 2½ months the recon project was underway.)

The serials conversion project just described was precipitated by our desire to have our physical sciences serial records available online in LS/2000 in a timely manner. Subsequent to the completion of this project, and armed with the experience it provided, the library considered what other categories of materials should be converted. Various collections and subject groupings were considered, and the library decided to proceed with a full-scale conversion of the rest of the serials collection, encompassing the holdings of the main library and the music library.

The reasons for converting the rest of the serials collection were somewhat more complex than the timing factors that influenced our decision to convert the physical and biological sciences collections. Essentially there were three main considerations. The first was our desire to have the bulk of our serials collection, which had virtually no OCLC records except the physical and biological sciences conversion, converted by qualified staff in a reasonable time frame. We felt that our serials cataloging staff would be better used in resolving problems identified during the OCLC conversion than in trying to do the huge conversion themselves. And, of course, they still had to keep up with the normal cataloging of new titles, so their time was limited.

We also hoped that if all of our serials were converted and loaded into LS/2000 we could start relying on the online catalog and stop costly and time-consuming production and filing of catalog cards.

Further, we hoped to be able to take advantage of the OCLC SC350 serials system, which at that time (summer 1984) had been announced for January 1985 availability. We envisioned online check-in and abandonment of the double input required by having separate serials and cataloging systems.

As it turned out, not all of these hopes were fulfilled. We are still producing catalog cards for serials because of delays in loading the LS/2000 database. We have, however, stopped producing any revised catalog cards for the physical sciences library collection, and eventually new catalog cards for the physical sciences collection will be filed only in the main union catalog. The SC350 serials system took longer to develop than we had anticipated, and its ability to accommodate a large serials collection led us to take a cautious approach; thus, we still do not have an online serials system.

However, from May 15, 1985, to December 30, 1985, OCLC converted over 21,000 additional serial records for us, a feat we could never have accomplished in that time. An additional 1,142 records were identified as problems and not converted, and 243 were found to have been converted in-house during the course of the second conversion project. Approximately 94 percent of the serials actually sent to OCLC for conversion were converted by the OCLC conversion team during the course of the two projects. Five percent of the batch were returned for in-house conversion, and 1 percent had already been updated during the course of the projects.

RESULTS OF THE CONVERSION

OCLC returned our serials shelflist cards in call number order. All records that had been converted with no problems were grouped in the front of the boxes, and those records that had been flagged for one reason or another were grouped in the back of the box. Clerical staff took the first batch of shelflist cards and refiled them into the main library shelflist, pulling and discarding the temporary shelflist slips. The second batch of shelflists, those that had been flagged, were then sorted into three groups.

The first group included records converted by OCLC, but in these records OCLC had noted names that differed from the online record. Clerical staff took these shelflist records, photocopied them, made a printout of the online record, and then refiled the shelflist cards. The photocopy of the shelflist and printouts were then used to create a card file in which the nature of the authority problem was noted. The majority of the serials authority discrepancies involved cases in which our catalog cards had pre-*AACR2* forms of names and the online record had an *AACR2* form of name. The reverse situation is also possible, however, where our record had an *AACR2* form of name and the online record had a pre-*AACR2* form of name. These authority problems will be addressed, as time permits, and eventually all entries will be in *AACR2* form for LS/2000.

The second group of shelflist cards returned by the OCLC team (about 2,000 records) was also converted, but OCLC had questions about one or more fields and therefore set them aside for us to verify. The last group included the 1,376 shelflist cards of records not converted by OCLC due to cataloging problems. These last two groups are still being verified or converted in-house.

OCLC also flagged other converted records, with minor discrepancies noted, that required additional scrutiny. If those converted titles are included, a total of about 3,776 serials titles require further attention. As of August 1987 approximately 1,000 titles have been examined by our serials cataloging staff. The nature of the problems revealed is discussed below.

Our original idea was not to create any new cards when converting or verifying records (or no more than absolutely necessary) but only to correct the OCLC holdings and LS/2000 records. This decision was made in consultation with the public services division of the library. We have, however, had to create many more card sets than anticipated to get a more complete or accurate shelflist card. In some cases the already existing latest entry shelflist card did not have an adequate outline of the history of title changes associated with a serial and a fuller accounting of those changes was deemed necessary. In other cases, the complications of forms of entry, errors in chronological and numerical data, or problems with other related data were such that it was easier to produce a new shelflist by creating a full set of catalog cards than to correct the existing one. Actually, once all of the necessary editing of the OCLC copy for conversion of such a title had been done, producing a set of new cards added little extra work.

An additional problem in the in-house conversion of records set aside by OCLC has been, and is, the authority work involved. Our ultimate goal is to update all entries to *AACR2* form in the LS/2000 system. Since the library is committed to maintaining the main library's public card catalog until the LS/2000 system is up and running for a large percentage of the library's records, this involves creating links between *AACR2* and pre-*AACR2* records for the main library's public card catalog as well as in the LS/2000 system. Therefore, we are checking all name and series entries for each of those records set aside by OCLC as having authority problems. We are also updating those entries to the *AACR2* form and creating links between names when necessary. We are maintaining these records in our manual name authority file for now. Eventually, discrepancies in forms of names, as noted in the name authority file, will be corrected in LS/2000 when we have the capability of making global authority changes.

The types of serials problems we have encountered in converting the records that OCLC was not able to convert fall into two categories.

One category of problems was latest entry records that were not straightforward title splits. For example, there were cases where the OCLC database did not include any or all of the serial titles represented in our latest title entry records. Although our contract required OCLC to enter new records for such cases, sometimes our catalog cards did not provide enough information for them to do so. In resolving these problems we have had to create many original successive entries. There were also discrepancies between our chronological and numerical data and those of the online record that prevented OCLC from splitting a latest entry card. OCLC set those records aside for us to resolve, and to do so it was necessary to examine the actual volumes. Quite often we found our data to be correct.

Another category of problems set aside by OCLC for us to resolve was UMass serials cataloging for which no corresponding serial record had been found: instead OCLC found monographic cataloging created by the Library of Congress or another member library. We have found this problem with some of our legal materials. In the past we have tended to catalog certain types of publications, such as loose-leafs and other updated items, as serials. We did this because we wanted the convenience of checking in updates in our central Kardex. Now the Library of Congress has published clearer guidelines for this type of material and catalogs them as monographs. We have recataloged these items as monographs, but we still include holdings information in the serials database and check them in, as appropriate, in our central Kardex.

In addition to converting in-house special problems set aside by OCLC, we are in the process of verifying the accuracy of information in some of the converted records. The problems we have uncovered so far are as follows:

For some records there is a discrepancy between our catalog record and the online record concerning the volume or date information. Again, resolving this problem involved looking at the piece itself and deciding which record was right. For other records the title (245 field) on our catalog card does not exactly match the title (245 field) of the online record but matches another field in the online record. For example, our title proper might match another title in the 246 field of the online record. A good example of this is an initialism that we have recorded as the title proper on our catalog cards but that is spelled out in full in the online record. (In this case both titles would have appeared on the piece, and there was uncertainty as to which was the true title.) So, although OCLC had identified the online record as the correct match and converted our title to that record, it was set aside for us to examine and decide if we wanted to make any changes for our LS/2000 record.

In other records converted by OCLC, differences between the links with other titles (780 and 785 fields) on our records and the online record were noted and left for us to decide which were correct. In some cases OCLC set aside converted records because of possible errors on our cards, such as an inverted digit in a date. A number of these records have been examined and simply refiled in the shelflist with no corrections needed. Others have had to be recataloged.

CONCLUSION

Final conclusions concerning the success of this project are still being formulated; more use of the records by library staff and patrons is needed. However, the creation of individual item records and linking the physical sciences library's catalog records to them in the LS/2000 system was completed in fall 1986. That library has operated on the LS/2000 system since January 1987 without any major problems from the conversion being detected. Based on these results, our reaction to the conversion project is a positive one. Over 27,000 serial titles were converted in a relatively short period of time, much shorter than if we had done them all in-house.

Also, problem serials and the nature of those problems were identified for us by the OCLC conversion team, which has saved us time in our in-house analysis and conversion. We are confident that the serials retrospective conversion project has achieved its goal of a timely and high-quality conversion of a large number of records and trust that future use of these records will justify this confidence. ■■

Twenty Years Ago in JOLA

It was the policy at the Yale Library for all programming to be done by library programmers, since various inefficiencies, and indeed catastrophes, had occasionally been observed when non-library personnel had prepared programs for library operations. The single exception to this policy was the proof program, which this investigation reveals used an exhorbitant amount of time—one-third of that required for subsequent card production. Since it had been felt that writing and coding a proof-listing program was perfectly straightforward, an outside programmer of recognized ability was employed to write and code the program. Because the program was simple, and because the programmer had high competence, efficiency of the program was never checked as it should have been.

This episode raises the question that if even the wary can be trapped, how can the unwary avoid pitfalls? There is no satisfactory answer, but it would appear that some difficulties could be avoided by review of new programs by experienced library programmers, of which there are unfortunately far too few. . . . Of course, when widely used library computer programs of recognized efficiency are generally available, magnitude of the pitfalls will have been greatly reduced.

Frederick G. Kilgour, "Costs of Library Catalog Cards Produced by Computer," *JOLA* 1, no.2:126 (June 1968).

Management Information Gleaned from Automated Library Systems

Carol Pitts Hawks

One of the functions of automated library systems is to provide management information to aid in decision making. The increase in size and complexity of library operations means that managers need an effective system to supplement information gained through direct observation. Management information generated by automated library systems in collection development, acquisitions and serials, cataloging, the online catalog, and circulation is discussed. Specific examples and their use are provided.

Edwin Cortez has defined an automated library system by its capacity to store "the library's files in a common database which is accessible through a common protocol and command language. The integrated system supports circulation, acquisitions, and cataloging activities, and for management's use generates reports for each of these activities."[1] Most automated library systems are designed with a dual function in mind—to support day-to-day library operations and to provide managers with useful information to aid in decision making. Therefore, systems must be evaluated not only for how effectively they handle day-to-day operations but also for their ability to manipulate and generate information for management.[2] It is this last function—management information—that will be the focus of this article.

Management and the consequent need for management information exists at all levels of policy and responsibility within an organization. The services offered by libraries have increased dramatically with the advent of technology. The resulting increase in the size and complexity of library operations has meant that most managers need an effective system to supplement information gained through direct observation. There can be little doubt that the more one knows, the better one can manage. The essence of good management is good information. In this sense, management as a function can be defined as interpretation of information so that appropriate decisions can be made.[3]

It is important to stress, however, the distinction between data and information. Data are nothing more than series of facts extracted for use by an individual. Information is the meaning derived from data by an individual.[4] Computers can generate reams of printouts containing massive detail. Such excesses can work against, rather than in favor of, competent decision making.[5] It is important to ignore transactions that fall within normal parameters and concentrate, instead, on the abnormalities and determine the reason for their deviance. These exceptions can point to problems in work flow or procedures that can be

Carol Pitts Hawks is Head, Acquisitions Department, Ohio State University Libraries, Columbus. A version of the paper was presented at the Sixth Texas Conference on Library Automation, October 1987, in Houston.

simply rectified before the problems get out of control. In other words, internal problems can be identified and corrected before the public is affected by them. Information overload can occur if computers are used carelessly, but when used appropriately, computers can process large sums of data rapidly and focus attention on items of concern. At the same time, it is important to be sure that the cost of collecting and analyzing data does not exceed the information's usefulness.[6]

The types of reports generated by automated library systems fall into four broad categories: statistical, exception, on-demand, and predictive. Computers are well equipped to provide this information because they were originally designed to be, and still primarily are, number crunchers.

STATISTICAL REPORTS

Manually derived statistics are labor-intensive and subject to error. They often lack sufficient currency and completeness to be truly useful. Two obvious advantages of using automation to compile statistics are the increased reliability of the numbers and the decreased staff time required to compile them. Automation permits analysis of operations never before feasible. Periodic reports provide routine, statistical information in detailed or summarized form. Not only can activities or transactions be counted, but they can be aggregated by time period and/or location, for example, and by numbers of items acquired, circulated, or cataloged. Statistical reports can also provide information based upon such factors as the time to acquire an item, to catalog materials, or to complete an interlibrary loan. Automation can also generate statistical reports in the form of ratios and percentages, such as the percent of registered borrowers who use the library monthly.[7]

EXCEPTION REPORTS

Exception reporting has been mentioned earlier in terms of locating items outside of normal parameters. Exception reports highlight areas requiring managerial attention. For example, a list of *all* books sent to cataloging would not focus on those that have been overlooked or lost. The overdue in cataloging report, on the other hand, provides the exact focus required. Only those titles not cataloged within a given time period are reported. The Geac system is built upon exception reporting. Most functions within the system operate on a series of action dates. Once this action date has passed without the appropriate action being taken, the item appears on the relevant exception report. The item will remain on the report until the exception has been resolved. Therefore, reports can be reprinted on a periodic basis and will only report the exceptions not yet resolved.

ON-DEMAND REPORTS

On-demand reports provide a response to a particular nonstandard question. For example, such a query might be, How many books in the TX classification have circulated in the last semester? The common feature of on-demand reports is that they are produced in response to a particular inquiry and not on a regular basis. These reports often employ Boolean logic or are part of a piece of software known as a report generator. Report generators are becoming a common feature in integrated library systems. They enable the library to specify the data elements to be included in the report and the format or sequence to be used. The generators do not require a knowledge of programming, and it is almost impossible to damage the database or software since the program extracts, but does not update or modify, files in the database.

PREDICTIVE REPORTS

Predictive reports give forecasts and provide comparisons based on statistical manipulation of data. These reports provide information on how the library is doing as compared to previous years, what trends are emerging, where activity growth is occurring, or how the library relates to other similar libraries. For example, a library may know that interlibrary loan lending has increased 48 percent. What is unknown is whether this is the result of improved service (i.e., timeliness), institution of interlibrary loan fees by other libraries, or improved collection development. Although this type of information is of primary interest to high-level management, it also has applications at the middle-management level. For example, predictive reports can project work flows and predict slow periods when staff need to be reallocated. In addition, predictive reports can often assist managers with "what if" situations.[8] For example, if the materials budget were reduced 8 percent, how would that affect resources for serials and other materials? Historical data can help assess the impact this change might have on a library and its operation and the consequences of alternative actions. However, it is important to remember that the system never makes the decision; it only provides the framework and data within which the manager must make the decision.

AUTOMATED LIBRARY SYSTEMS

Library school professor Jean Tague states that

> the greatest potential for management information from automated systems lies in the control and planning of individual subsystems. Automated subsystems provide the capability for a kind of analysis of operations never before feasible by an over-worked library staff.[9]

Management information was often an afterthought in early automated systems. However, decision making has become such a critical part of everyday library management that systems are now being designed with management information components from the very beginning. Library management in the areas of collection development, acquisitions/serials, cataloging, the online catalog, and circulation can be greatly enhanced by the use of management information.

COLLECTION DEVELOPMENT

Although there is no collection development subsystem in any current integrated library system (ILS), collection development can be well served by the information available from the circulation and acquisitions modules. One of the goals of collection development is to provide materials in advance of requests. Therefore, one of the most important pieces of information in improving collection development is determining how much the existing collection is used. A University of Pittsburgh library study showed that only 56 percent of newly acquired materials actually circulated. Heaviest use occurred in the two years following acquisition; if a book has not circulated within five years of acquisition, it probably never will circulate. Obviously, more extensive analysis of circulation data is warranted in order to improve collection development.[10]

Several reports in most integrated library systems can address these issues. The first report covers the circulation statistics by patron types. This report generates information on the category of patrons using the library. These categories are usually defined by the individual library when the system is installed and can be extremely detailed if desired.

The second report—checkouts and renewals by call number range—compiles statistics on those ranges most heavily circulated. Such information is indicative of emerging trends and research interests. Depending upon the level of detail in the report, definite trends can be determined by comparing data from several semesters. Both of these reports can be used by collection development to redefine and reevaluate collection policies and goals.

Circulation systems can also provide useful information in the area of high-demand material. If books on a particular subject are in high demand and attract frequent holds, additional purchases may be justified. The in-house library system at Bell Laboratories produces a weekly "Titles in Demand List" to identify all titles for which five or more people are waiting. Supervisors then hold a weekly conference to determine additional purchases necessary to meet demands and keep response time short. The Geac system has a similar report—the Holds/Purchase Alert Report—which allows each library to set the number of holds that will trigger the alert. Full bibliographic information is generated and forwarded to collection development for a decision.

At the opposite extreme, systems can generate lists such as Bell Labs' "Zero Activity Report" to identify titles that have little or no use in a given time period. In addition to the time period, location and subject area can also be limited on such reports. Unusually low use may call for increased user education or more publicity. On the other hand, as more libraries face weeding and storage decisions, such an analysis of circulation data can enhance those decisions. Unfortunately, some systems cannot report items that have never circulated.

ACQUISITIONS/SERIALS

The term *acquisitions* encompasses every aspect of obtaining materials for libraries. Materials can be acquired by purchase, gift, or exchange. Even though libraries approach this process from many perspectives, automation has greatly improved the service and efficiency of such operations.

Financial management is one area of acquisitions where automation has yielded marked improvements. Systems such as Geac and Innovacq recognize the library's need for double bookkeeping or hierarchical breakdowns within accounts. This allows libraries to report expenditures to their parent accounting bodies in general terms as well as to maintain detailed internal subject and format breakdowns. Not only are detailed breakdowns accommodated, but encumbrances and expenditures are updated in real time or—in the worst case—overnight, requiring no human tabulation. Based upon this capability, financial reports can then be generated in a format suitable for general circulation. In addition to basic information such as budget, expenditures, encumbrances, and free balances, most systems can tabulate such things as the percent encumbered/expended or a comparison between current and prior year expenditures.

Currency conversion has long been a problem for library accounting divisions. The estimated prices of titles can be entered in a foreign currency, and the system will automatically convert and encumber funds in U.S. dollars. As conversion rates fluctuate, the new rate can be entered as often as necessary. At that time, each outstanding encumbrance is recalculated at the new rate. The Geac system generates two management reports in the conversion process—one reports the plus/minus difference in each currency, and the other reports how the change has affected each individual fund.

Projecting the cost of serials consistently poses one of the most complex financial management problems. The rising cost of serials demands close managerial attention. Nationwide studies can provide an average inflation rate, but this average may be seriously inaccurate given a library's particular mix of serials. Automated systems enable libraries to record and maintain an invoicing history for multiple years. This invoicing history can report year-to-year inflation rates. The Geac system then uses a complicated weighing formula to project costs on a title-by-title basis. This weighing formula takes into account the latest subscription period and gives greater emphasis to the most recent payments. Recently, several major subscription vendors such as Faxon have also developed systems to provide this information. However, few libraries, if any, maintain all standing orders with a single vendor. Therefore, it is important that local systems be designed to provide this information. With in-house systems, records can be manipulated to report data in various combinations, such as by fund, by departments or college, or by geographic source of publication.

Vendor performance reports summarize average order costs, discount percentages, delivery times, and the number of claims and cancellations. The usefulness of such reports to management is in renewing purchase orders and maintaining effective supply sources. However, it would also be useful if systems were designed to report the vendors who exceed a particular delivery time rather than only producing detailed reports on how long each vendor takes to deliver the average order. This is an example of isolating the exceptions rather than reporting all of the normal parameters.

Order control is the final area of acquisitions where management information has increased and improved with the advent of automation and where exception reporting becomes particularly useful. NOTIS as well as many other systems can produce on-demand lists of all active orders. For example, titles that have not been received by their estimated delivery date can be extracted and investigated by staff. In some cases, automatic claims can be generated from such extractions. Systems such as NOTIS can compose letters for common conditions such as claims, cancellations, or credit requests. Terminal operators need only enter mnemonic codes, which the system explodes into full-text paragraphs. NOTIS can even print these messages in languages other than English. The Bell Labs' system generates the "In Process File Scoreboard," a report that permits a manager not only to determine how many items are in any given status, such as on order, but also to compare the current status with an earlier date, such as last month.[11] Such a comparison enables the manager to determine if progress has been made in eliminating backlogs. Acquisitions practices vary so widely in libraries that it is a tribute to automation vendors that recent acquisitions systems have become so successful in meeting library expectations for management information.

CATALOGING

Management reports available from the cataloging subsystem of an ILS are oriented primarily toward work-flow analysis. Records that do not meet certain standards are bounced from the process to be reviewed. Because of the complex flow of new cataloging, on-order records, and authority records, a sound series of management reports is required to manage the work flow.

One of the almost immediate uses of cataloging reports is in estimating the amount of time required for completing the initial database load. A university in the U.K. used its initial data entry statistics to predict completion of the task. Dates for making the system available to the public were then projected, and a public relations campaign planned.

> The fact that this operation could be constantly monitored and targets redefined in order to achieve success attests to the value in having detailed and accurate information, thereby enabling management to know exactly what progress was being made.[12]

Most cataloging databases are built by tape loading of OCLC or other machine-readable records. Once this initial load is accomplished, however, new cataloging or retrospective conversion titles are merged into the database through some type of interface. In order to maintain the integrity and quality of the database, transferred records are run through a series of checkers. Items not meeting all of the appropriate standards are printed in brief bibliographic form. This proof list (as it is called by Geac) uses an asterisk to identify the tag where the problem exists. Standard error codes are employed, and after corrections are completed the record will be sent back through the series of checkers. If no further errors are found, the record will be added to the database.

Beyond this, the successful day-to-day operations of processing work flow depend upon a number of management reports. Of particular use in managing the large, specialized subject headings authority database is a periodic printout of all headings used within a given class number. Two NOTIS reports—the "New Subject Headings List" and the "Dropped Subject Headings List"—provide cataloging staff with a way to maintain catalog integrity. Staff can use these lists to prompt postcataloging authority work. The ease with which authority work can be managed has been greatly enhanced by automation. Individual headings can be changed throughout the system in a global update with a minimum of effort. Lists of such changes can be used to notify member libraries in a network environment of how headings have changed.

THE ONLINE CATALOG

Catalogs are the key to library collections, and understanding their use could serve as the key to understanding library use. For the first time it is now possible to study, on a routine basis, how patrons use the catalog. The essential ingredient for catalog-use assessment is the ability to record searches automatically for subsequent analysis. Transaction logs are the standard mechanisms for recording this data. Such logs are records of the computer dialogue between the user and the catalog. Monitoring is achieved directly and unobtrusively. Because most catalogs are not passworded, personal identification is impossible, and the privacy of the individual is preserved.

Various statistics can be compiled through analysis of the transaction file. Because only transactions are recorded, reports do not show data from a period when a terminal has not been used. Statistics are subdivided by terminal, by hourly periods, and, within each hour, by type of search or activity. It is therefore possible to see how many searches, in a given hour, were performed on a particular terminal and whether they were searches by author, title, subject, etc. The use of function keys and major functions can also be analyzed. In one such analysis, the University of Sussex Library discovered that the popularity of a new search technique was rising. This particular search method could not be performed with a card catalog. Therefore, it is significant that about 70 percent of the time users chose a type of search that would not have been possible in the card catalog.[13]

Beyond the obvious advantages of learning more about library users, transaction logs can be used to improve the design of the online catalog. Improvements in screen layouts and filing sequences can be prompted by user studies. Summaries of transactions provide information useful in deciding on terminal placement, improving screen effectiveness, and determining the need for cross-references. In a recent book on the Geac system, Duncan Westlake succinctly states that transaction logs are

> another example of the way in which an automated system can be made to produce genuinely relevant statistics, with the possibility that decision making can be made more reliable in the light of better information.[14]

CIRCULATION

Circulation control is concerned with issuing, returning, and reserving materials and identifying overdue items. The software designed for circulation must provide for the creation of printed output and financial control. It is an especially rich source of management information. The information available relates to individual books, categories of items, categories of readers, and transaction activity.

As mentioned in other sections, work flow and staffing analysis are natural byproducts of automated systems. Most systems produce a daily operations report providing a history of all transactions of the previous day and statistics for the various circulation activities. The report provides information for any follow-up work the staff needs to do and enables them to monitor circulation activity. This allows a library to analyze busy times, since most reports are subdivided by hourly periods. For example, a daily operations report has potential for dynamic staffing of small branch libraries by allocating staff to cope with anticipated peak periods. Also, work that tends to degrade the response time of the system can be scheduled during slack periods when the system is not heavily used.

Borrower activity can be scrutinized in much the same way as catalog use. The Geac system provides for patron extract programs, which function using Boolean operators. The library can define a number of criteria, then either list or count the number of patrons who meet them. Information that can be defined includes such items as whether the borrower's privileges have been suspended, whether the borrower has overdue books, and whether the borrower has an incomplete address. Combining some of these items results in reports comparing such things as active versus inactive borrowers or which patrons registered at one branch are borrowing extensively from another.

Most systems now have a variety of programs dealing with hold management and class reserve capabilities. A *hold* is defined as the process by which a patron is notified that a requested item is available and can be picked up at the library. In addition to reporting the number of holds placed, systems can report the number of items not picked up. This information can be analyzed to determine which patrons are not picking up holds in order to see if there are notification or other problems with the process. In addition, the library can keep track of the number of holds not satisfied.

In an academic setting, reserves are an integral component of a circulation system. Geac's report, the "Weeding and Stamping List," automatically places on or removes items from reserve. This is a powerful program that operates on activation and expiration dates to provide automatic updating. A list of the actual items can be printed, which is then used to pull items to be placed in the reserve room or as a listing of what is currently on reserve.

A second report, the "Reserve Use Report," is arranged by course number and professor's name. It is designed to give professors information on which items placed on reserve are used and which are not. The circulation count as well as the reshelve count is reported. These can be presented to faculty to encourage them to consider the usefulness of some of the titles on reserve.

Automated circulation systems allow libraries to track material more closely, particularly those items declared missing. When an item is declared missing, it can be updated in the system. Lists can be generated to facilitate searching for these titles. The reason for the missing status is included as well as the length of time the item has been lost. Both NOTIS and Geac notify through reports when missing items have reappeared. In such a case, a refund may be warranted. A list of patrons eligible for refunds can be produced. This permits the library to make a refund when a book paid for by a patron simply shows up.

After a certain length of time, items may be declared permanently lost. They can be purged from the circulation system and then forwarded to cataloging for permanent withdrawal. The list could then be sent on to the collection development staff for replacement decisions if this was not done at an earlier stage.

Fine policies vary widely among libraries, but all libraries must contend with the problem of overdue material. At the very least, overdue and recall notices can be generated through most systems. The system should also provide for detailed information concerning fines or replacement bills. In addition, the library must be able to block patron usage through the system when financial obligations reach certain levels. More and more often, libraries are including specifications in bid documents requiring programs to transfer financial information from the system to the university business office. This linking can be extremely beneficial for libraries. After implementing such a process, the University of Houston's success rate at retrieving long overdue items or replacement fees increased significantly.

CONCLUSION

An important aspect of any automated library system is its ability to produce information about how the system is being used. Much of the enthusiasm for automated systems comes from the power and responsiveness of the systems in providing data. Automation makes large quantities of raw data almost instantaneously available to managers. Interconnected systems are also able to realize relationships between information that might not be apparent in manual systems. It is important to bear in mind that a primary purpose of automation is to assist library management in allocating resources, selecting alternatives, and providing timely information. To work effectively, the manager's information needs must be well thought out and documented so that relevant information is generated.

REFERENCES

1. Edwin M. Cortez, "Library Automation and Management Information Systems," *Journal of Library Administration* 4, no.3:28 (Fall 1983).
2. Ibid. p.22–24.
3. J. Michael Bruer, "Management Information Aspects of Automated Acquisitions Systems," *Library Resources and Technical Services* 24, no.4:339 (Fall 1980).
4. Richard J. Boland, Jr., "Tutorial on Management Information Systems," in Clinic on Library Applications in Data Processing, University of Illiniois, *Proceedings: Library Automation as a Source of Management Information* (Urbana–Champaign, Ill.: Univ. of Illinois, Graduate School of Library and Information Science, 1982), p.10.
5. Linda Main, "Decision Support with Decision-Modeling Software," *Library Software Review* 6, no.3:128 (May–June 1987).
6. Mary Ellen Jacob, "Management Information Systems in a Network Environment," in Clinic on Library Applications in Data Processing, University of Illinois, *Proceedings: Library Automation as a Source of Management Information* (Urbana–Champaign, Ill.: Univ. of Illinois, Graduate School of Library and Information Science, 1982), p.112–14.
7. Jean Tague, "Computer Potential for Man-

agement Information," *Canadian Library Journal* 35, no.5:269 (Oct. 1979).
8. Jacob, "Management Information Systems," p.116–17.
9. Tague, "Computer Potential," p.269.
10. Richard Boss, "Automating Acquisitions," *Library Technology Reports* 22:499, 509 (Sept.–Oct. 1986).
11. Robert A. Kennedy, "Computer-Derived Management Information in a Special Library," in Clinic on Library Applications in Data Processing, University of Illinois, *Proceedings: Library Automation as a Source of Management Information* (Urbana-Champaign, Ill.: Univ. of Illinois, Graduate School of Library and Information Science, 1982), p.136.
12. Duncan R. Westlake, *Geac: A Guide for Librarians and Systems Managers* (Brookfield, Vt.: Gower, 1987), p.242.
13. R. C. Young, "The Geac-Based OPAC at the University of Sussex Library," *Program* 20, no.2:146–47 (Apr. 1986).
14. Westlake, *Geac: A Guide*, p.245–47.

BIBLIOGRAPHY

Boss, Richard. "Automating Acquisitions." *Library Technology Reports* 22:479–637 (Sept.–Oct. 1986).

Bruer, J. Michael. "Management Information Aspects of Automated Acquisitions Systems." *Library Resources and Technical Services* 24, no.4:339–42 (Fall 1980).

Clinic on Library Applications in Data Processing, University of Illinois. *Proceedings: Library Automation as a Source of Management Information*. Urbana-Champaign, Ill.: Univ. of Illinois, Graduate School of Library and Information Science, 1982.

Cortez, Edwin M. "Library Automation and Management Information System." *Journal of Library Administration* 4, no.3:21–33 (Fall 1983).

Desmarais, Norman. "BataSystems Acquisitions." *Library Software Review* 6, no.3:146–49 (May–June 1987).

________. "Sydney Micro Library System." *Library Software Review* 5, no.5:306–15 (Sept.–Oct. 1986).

Hawks, Carol P. "The Geac Acquisitions System as a Source of Management Information." *Library Acquisitions: Practice & Theory* 10, no.4:245–53 (1986).

Herceg, Joyce D. "Acquisitions Management Information on a Shoestring." *Small Computers in Libraries* 6, no.8:20–28 (Sept. 1986).

Leggate, Peter. "The Microcomputer in the Library: IV. Cataloguing and Acquisitions Software." *The Electronic Library* 4, no.3:152–65 (June 1986).

Leggate, Peter. "The Microcomputer in the Library: V. Circulation Control and Serials Control." *The Electronic Library* 4, no.4:218–29 (Aug. 1986).

Magnuson, Barbara. "Collection Management: New Technology, New Dimensions." *Wilson Library Bulletin* 57, no.9:736–41 (May 1983).

Main, Linda. "Decision Support with Decision-Modeling Software." *Library Software Review* 6, no.3:128–33 (May–June 1987).

Moskowitz, Mickey. "The Collection Manager's Micro." *Technicalities* 5, no.3:11–14 (Mar. 1985).

Perry-Holmes, Claudia. "Lotus 1-2-3 and Decision Support: Allocating the Monograph Budget." *Library Software Review* 4, no.4:205–13 (July–Aug. 1985).

Rush, James E. "Automated Serials Control Systems." *Serials Review* 12, no.2–3:87–101 (Summer/Fall 1986).

Tague, Jean. "Computer Potential for Management Information." *Canadian Library Journal* 36, no.5:268–70 (Oct. 1979).

Westlake, Duncan R. *Geac: A Guide for Librarians and Systems Managers*. Brookfield, Vt.: Gower, 1987.

Young, R. C. "The Geac-Based OPAC at the University of Sussex Library." *Program* 20, no.2:138–50 (Apr. 1986). ■■

The State of Authority

Nadine L. Baer and Karl E. Johnson

Surveys were conducted to determine how American college and university libraries are employing authority control, the impact of AACR2 and online catalogs upon the need for authority control, and the effect of OCLC cataloging upon a university library catalog established without authority control for corporate or personal names. The conclusions were that most American college and university libraries utilize authority control over most entries, AACR2 and online catalogs are responsible for an increased interest in authority control, the trend toward online catalogs with linked authority control will greatly enhance control over cataloging records, and cataloging from the OCLC database can be expected to conflict significantly in a library catalog without authority control.

The concept of authority control over library catalogs has been with us since the time of Charles Ammi Cutter. Few library school students graduate without some understanding of the need for authority control over library collections. Rare is the working cataloger without a strong desire to establish but one "authoritative" form for each given name entry with effective linkages to and from related entry terms. Yet, authority control often is perceived as less important than getting the books onto the shelves quickly.

The authors have served for a combined total of more than fifty years in a university library setting where authority control (for all except subjects and, more recently, series) has been seen as an unattainable dream, due primarily to underfunding resulting in understaffing and lack of space and equipment. It is within this framework that the authors decided to investigate the utilization of authority control in American college and university libraries to determine if other libraries face similar problems and if effective authority control is in fact commonly unattainable.

Three surveys were conducted during the course of the investigation. They included a survey of the literature, a survey of how American college and university libraries employ authority control, and a survey of corporate and personal name entries in the author/title catalog at the University of Rhode Island main library.

SURVEY OF THE LITERATURE

A literature survey restricted largely to *Library Literature* from 1974 to the present produced some seventy references pertaining to authority control. The literature was examined with a tentative assumption that, with time, evidence might surface that would make a case in favor of lessened authority control as more advanced computerized searching systems became available. This did not turn out to be the case, as the great majority of publications recognized the need for extensive authority control

---

Nadine L. Baer and Karl E. Johnson are both with the Catalog Department, University of Rhode Island Library, Kingston.

over library catalogs, regardless of whether or not the library catalog was in machine-readable form. Some research[1,2] did suggest that authority control is somewhat less important in an online catalog equipped with keyword indexing and right-hand truncation, but no authors advocated total abandonment of authority control; Frederick G. Kilgour of OCLC, however, did see authority control "as going backwards toward the 19th century instead of going forward in a catalog design that doesn't require that kind of control."[3] Brief summaries of the more relevant publications, depicting the chronology of the historical development of the need for effective authority control, follow.

As early as 1972, Grosch[4] recognized the need for authority control and discussed implementation of computer-based subject authority files at the University of Minnesota libraries. Grosch also noted that little had been done prior to this time by systems designers to develop automated subject control. Ohmes and Jones,[5] in a 1973 evaluation of the Ohio College Library Center (OCLC) bibliographic retrieval/card production system, perceived that the OCLC system, due to its inclusion of cataloging records produced over a period of time without concern for reconciliation of old and new forms of headings (i.e., lack of authority control), failed "to be of help to the cataloger in providing an expeditious way to make the catalog consistent." Malinconico,[6] in his 1974 paper describing automated authority control at the New York Public Library, stated

> It should be obvious that the library catalog is considerably more than another manifestation of simple document retrieval. Representing the relationship of one item to others in the collection is perhaps more important than providing access to a single item.

Malinconico concluded by stating

> I believe the system we have developed at NYPL, thus far, should be ample proof of the feasibility of authority control in practice. I would also hope that I have managed to convey the importance of such a mechanism in maintaining the catalog of an individual library. Mechanical control over the data in a data base collected from several disparate sources, as in the collection of a national data base, is of even greater importance. Only in this way can we insure that we will be collecting a consistent coherent data base and not a chaotic machine readable collection of reference records.

Malinconico,[7] in a 1976 discussion of computerized library catalogs, stated that "a catalog . . . should manifest . . . a coherent, logically consistent structure" that "must be imposed from the outset" by the cataloger. Malinconico also noted that "the machine is indeed quite adept at creating alternative access points . . . but . . . only on data in the form in which it receives it. . . . It cannot impose an order that is not already there." Atkinson,[8] in his 1979 description of the electronic catalog used at Ohio State University, noted that it was a system for known-item searching only and declared, "We do not propose to make a complete substitution of an electronic system for the card catalog." He stated that "the serious problems are not the cross-reference problems, the syndetic structures, or the technical complications of authority files" and went on to say that

> we have had great success using an algorithm or a code technique of searching the rather slim records we use with the . . . system . . . [and] lack of precision has not proved to be a detriment at all . . . since the shortened or truncated search code demands far less accuracy than the longer or more complete ones.

Bierman,[9] in 1979, predicted that "beyond 1985, online catalogs containing authority-quality control systems will increasingly become the single form of catalog in use by an increasing number of libraries of all types and sizes through regional-national-international interconnected networks." Kilgour,[10] in his 1979 presentation regarding online catalog design, described "a new catalog designed specifically for the user," one which "should be viewed as a huge number of miniature catalogs" and stated that

> from the viewpoint of the collection being cataloged, there is first the author-type miniature catalog which is analogous to an AACR main-entry catalog except that it is more complete. Second, there is the online title catalog, also similar to AACR, except that the online catalog is complete; it contains every title in the collection. Third, there is an author-title catalog for which there is no comparative AACR catalog.

Kilgour then went on to say that from the viewpoint of the online catalog designer, due to "the multiple retrieval power of an online catalog, the concept of main entry is not useful" and then pointed out the fact that "an online catalog contains many more retrieval accesses than a catalog based on the Paris Principles and is therefore a far more effective tool from the user's point of view." Lipow,[11] in a 1979 presentation of the alternatives to research library catalogs, indicated her concern "about the form which one would use to look for people, places, organizations, etc." and stated, "In my opinion, it is highly desirable to have consistency in form." She further noted that lack of authority work was due to budgetary constraints and "was recognized to be a step backward in service." Malinconico,[12] in a 1979 article on database organization and authority control, took exception to those who believe authority control to be a precomputer anachronism and concluded that "authority control of a cataloging data base is essential if that base is to support an effective machine readable catalog." Martin,[13] also in 1979, recognized the existence of "millions of dirty . . . or nonstandard bibliographical records," the result of lack of concern with authority control, and she made a plea for all libraries inputting data to the bibliographic utilities to use the MARC format and LC's interpretations of the Anglo-American cataloging rules. In 1980, a panel session of the American Association of Law Libraries' Seventy-Third Annual Meeting, entitled "Authority Control, or the Key to Survival in the Eighties,"[14] addressed the employment of authority control by the various bibliographic utilities and acknowledged the need of law libraries for authority control and the concern for the cost of individual systems. Matthews,[15] in her 1980 article on authority control, recognized that "historically authority control has not been a pressing issue in library management" and concluded that "authority files are necessary in any library for the sanity of the cataloger and the convenience of the patron . . . and need to be given more attention by libraries." Mosey,[16] in her 1980 dissertation examining how name authority control had changed as a result of catalog departments being able to produce cataloging via the OCLC utility, found that name authority control procedures in member libraries had not changed and stated

> The practices and procedures followed in name authority work were perhaps the most disappointing areas in the research. The cooperative spirit that a system like OCLC ideally engenders was not as prominent as expected. The practice of continuing to check contributed cataloging records for correct forms of names is both economically unsound and labor-intensive. The existence of an authority control subsystem would undoubtedly result in an increased level of confidence in other libraries' cataloging and lessen the need to be so self-reliant.

Runkle,[17] in a 1980 discussion of authority control and its relationship to online catalogs, stated that "it was OCLC that spread the advantages of shared cataloging throughout the country and made librarians more conscious of descriptive cataloging rules and standards" and further noted that "the question of the necessity of authority control in on-line catalogs for individual libraries may be passé. There are complications, however, related to automated authority control in shared bibliographic files." Schmierer,[18] in her 1980 article that examines the relationship between authority control and the library catalog, concluded that "the gathering function of the library catalog can be provided unambiguously only if authority control is provided" and went on to chide those "libraries that maintain files referred to as 'catalogs,' but which are not because for want of authority control these files do not provide the gathering function of the library catalog." Freedman,[19] in his 1981 interview of Fredrick G. Kilgour of OCLC, elicited from Kilgour the statement that OCLC was "developing an authority control system that will do a lot more than just authority control." Kilgour continued: "I'm not enthusiastic about authority control because I view it as going backwards toward the 19th century instead of going forward in a catalog design that doesn't require that kind of control." Kilgour also pointed out that, in the OCLC record, the use of the 87 × fields for different forms of a name served the purpose of cross-references. Bregzis,[20] in his 1982 comparison of Cutter's record syndetic catalog with the current entry syndetic catalog, suggested that "computer technology has given us an opportunity to return to the record syndetic structure of the catalog" and stated, "The automated method of record-to-record connection holds out the promise for some economy through a simpler management of bibliographic records." Bulaong,[21] in a 1982 discussion of authority control and cataloging standards, lamented the fact "that the major source agencies (except . . . Canada . . .) had no internal or comprehensive automated authority systems when bibliographic databases were being developed" and then strongly advocated authority control in both manual and automated environments, concluding that

> behind all this is an assumption that authorities are crucial to the collocation function of the catalog and that that function has to be maintained. There is a gnawing doubt whether they would be as crucial in a totally interactive and on-line environment, but the efficiency of searching and location in an on-line environment is directly enhanced by authority control. As information databases have shown, it is possible to get around the lack of authority control by the use of keywords, truncated and combined search keys or qualified searches using Boolean operators. But what we are striving for is exhaustiveness. In searching information databases, nobody knows what is being missed, because of lack of authority control.

Craig,[22] in his 1982 account of the misfortunes encountered by Glasgow University in its fourteen-year struggle to maintain consistency in series authority amid its participation in automation and cooperative cataloging, imparted a grim warning to any library considering uncontrolled use of cataloging records produced for a bibliographic utility lacking authority control. Culp,[23] in her 1982 description of authority control in the Washington Library Network, commented that "it is becoming apparent that, as a national database of contributed bibliographic information and holdings information develops, authority considerations are gaining importance. Only when disparate files are under bibliographic control can they be utilized and shared to their fullest potential." Dowell,[24] in a 1982 analysis of the impact of AACR2 headings on library catalogs, determined that between 14% and 20% of new headings would conflict with headings established under earlier rules, clearly more than ample justification for authority control, and concluded that "until library use of technology has made it possible to correct headings easily and without extensive error, the impact of new rules on existing catalogs should not be ignored entirely." Gorman,[25] also in 1982, discussed authority control in the catalog of the future and envisioned an automated authority control system "which will allow access to information by any form of a person's name or by any form of the name of a work which is present in an authority record"; i.e., even "the 'incorrect' forms will lead to the bibliographic records relating to the desired work." Ludy and Logan,[26] in their 1982 discussion of a machine-readable authority file linked to the Ohio State University online catalog, described the advantages of "the links between and among headings and bibliographic records" that not only "make automated authority control possible" but also provide for changes in headings, which are then "reflected in all bibliographic records containing that heading." Madden,[27] in her 1982 paper assessing vendors who provide authority control systems, discovered that "the many different vendors in the library marketplace who are concerned with authority control have vastly different views on the importance of authority control and very different priorities for their future development activities" and advised "those interested in authority control" to "speak loudly and frequently of your desire for authority control" and "investigate carefully what service you are getting and at what price." Matson,[28] in her 1982 article discussing the need for a "National Series Authority File," stated that "those of us who work intimately with series decisions are highly interested in the development of a nationwide series authority system. We hope that it will substantially improve our ability to maintain self-consistent local systems compatible with national practice." Miller,[29] in his 1982 paper on network authority control, noted that "in response to a questionnaire regarding the future plans of the AMIGOS Bibliographic Council, the network libraries stated as a very high priority that they wanted some form of network coordinated authority control" and further stated, "Many automated bibliographic management systems that are now being developed are considering authority management as the foundation for those systems." Perreault,[30] in a discourse on authority control, concluded that even in an online authority file "a controlled terminology is still needed, whether it is manifested in authorization/suppression (see-references) or in a network of equivalencies (see also-references)." Rather,[31] also in 1982, described authority control maintained by the Library of Congress (LC) and noted that "to LC, the backbone of effective bibliographic control is well-controlled authority data. It appears likely that it will also be a requirement in an effective national bibliographic system." Williamson,[32] in a 1982 look into the future, envisioned that in the library catalog of the year 2006, "authority files could assume an increasingly important role as control mechanisms, although their format and content may be significantly different from the authority files with which we are currently familiar." Baker,[33] in her 1983 article on the use of name authority in ten academic libraries in Arkansas, found that "nine libraries . . . were . . . involved in name-authority work to some degree and they did on-line cataloging through OCLC." Elias and Fair,[34] in their 1983 article describing a corporate experience with a computerized office communications system, found that "because of the inherently creative nature of individuals, a single staff member would often index the same entry in several different ways . . . for without any authority control a specific or correct form of the name did not exist." They concluded that "there always will be a need for authority control, whether filing is done in a card catalog or on a computer terminal." Funabiki, Mifflin, and Corlee,[35] in their 1983 description of the use of the Washington Library Network at Washington State University, concluded

> The authority control system of the Washington Library Network has been demonstrated to be a reliable, powerful, and versatile tool, resulting in efficient and comprehensive heading changes for Washington State University's research library. The cross-reference structure inherent in the authority control system has also resulted in an improved library catalog, which in turn provides better patron access to the materials in the Washington State University Libraries.

Henderson,[36] in a 1983 article on the need for authority control in computerized catalogs and databases, noted that changes in names of entries entering a library catalog are a reality and that

> while some persons feel that key word searching, truncated search keys, and other devices have eliminated the need for controlled vocabulary in on-line systems, others feel that computers have increased, not decreased, the need for authority systems. At present, many believe that the exchanging and merging of bibliographic data in machine-readable form is more easily accomplished if entries are compatible and if users can be assured that access points are uniformly represented.

She predicted, "For those who believe that authority control is the 'key to tomorrow's catalog' the next few years should see the development of many on-line authority control systems." Wajenberg,[37] in 1983, described how the University of Illinois at Urbana-Champaign (UIUC) was using the computer programs used in the Washington Library Network to develop an online catalog, incorporating authority control. Avram,[38] in her 1984 evaluation of authority control, concluded

> Some degree of authority control is required for the majority of library operations and it can be imposed anywhere in the network configuration or hierarchical structure. It follows that the work of imposing authority control can be performed once at a level of the network configuration at which many institutions are sharing the data, or the work can be performed by the many institutions themselves. The decision as to which way to go should be based on economics and we ought to reserve our energies for this analysis and stop questioning the validity of authority control.

Burger,[39] in a 1984 article relating the concepts of artificial intelligence to automated authority control in online catalogs, suggested that "machine based authority control is a form of artificial intelligence" already in use in the online catalogs of the Washington Library Network and the University of California. Dickson,[40] in her 1984 analysis of errors committed by patrons searching an online catalog, found that while many errors involved spelling, typing, or general ignorance of library catalog conventions (e.g., entry under surnames, omission of initial articles, etc.), there were instances where searches would have been helped by cross-references. Klemperer,[41] in a 1984 description of the use of authority control in the University of California online catalog, stated that "one of the most important features in the design of the MELVYL data base is the inclusion of a fully linked authority file." Liggett,[42] in a 1984 article describing the Name Authority Co-op (NACO) project at the Library of Congress, noted

> We see the development and growth of computer-facilitated cooperation in the major bibliographic utilities, but the lack of a common name authority file has been a continuing problem to their members. It is the view of the Library of Congress that we cannot enjoy the benefits of true bibliographic sharing on a national level until there is such a file.

Ludy and Rogers,[43] in their 1984 article on authority control at the Ohio State University libraries, found that for original cataloging, "the authority work done by the catalogers has not changed with automation. It is still done prior to cataloging and involves checking proposed headings against existing ones" and concluded that "authority work that is being done at the local library level is more complex, time-consuming, and costly because of the demands of resource sharing." Miller,[44] in his 1984 description of Blackwell North America's authority control system and its application to retrospective conversion of a library's records, noted

> Optimal return on the retrospective conversion investment includes the improved access that authority control provides, whether the file is used for COM fiche or film, or for an online catalog. Just as a misshelved book is a waste of the library's investment in purchasing the volume, so an incorrect heading that results in a patron missing a record is a waste of the library's investment in cataloging the item and in converting it to machine-readable form.

Rickerson,[45] in a 1984 description of the authority control system utilized by the Washington Library Network as well as future plans for the system, noted that "the considerable effort made at this time to provide support for authority control seems to indicate widespread interest." Shore,[46] in a 1984 comparison of personal names on title pages versus established headings, found that a difference existed in approximately 35% of the cases and concluded, "We can assume that the variation between use of names and the way headings are established is relevant to authority control issues." Taylor,[47] in a 1984 investigation of the value of authority files in online catalogs, determined that linked authority records would have helped in only 6% of unsuccessful personal name searches and concluded

> While I believe that the part of authority control that collocates all forms of a name under one form is essential, I also think that online linkage of all our elaborate authority records, the majority of which have no cross references, to a bibliographic file, which may contain only one record for each of many names in the authority file, is perhaps more expensive than is justifable.

Thomas,[48] in a 1984 study of the need for cross-references in the MELVYL online catalog at the University of California in San Diego, found that "47% of the cross-references catalogers now routinely make for manual files could be dispensed with for MELVYL," due to its use of "keyword indexing and automatic right-hand truncation." Burger,[49] in his 1985 book on authority control, stated

> Experienced catalogers are painfully aware that cataloging is ceaselessly changing. Codes change, local procedures change, LC rule interpretations change. The products of cataloging are forced to change because of these changes. Authority work is the process whereby librarians attempt to prevent the ceaselessly changing inputs and outputs of the cataloging process from creating bibliographic havoc.

Clack,[50] in her 1985 essay on authority control, noted

> Once the computer was introduced into the cataloging process, it became a common belief that all bibliographic retrieval problems would be solved. There was a misconception that integrity of the catalog was not necessary for an online catalog. . . . It was believed that with the computer a catalog could be produced that would be flexi-

ble and that could retrieve any information as long as it was in machine-readable form. . . . The computer does provide the online catalog with considerable flexibility; however, flexibility without the type of integrity generated by authority control does not make for a very efficient file

and concluded that "interest in authority control is on the rise, primarily, because of experiences with online catalogs." Epstein,[51] in her 1985 evaluation of automated authority control, pointed out many of the problems inherent in existing systems and recommended "that an online catalog be user friendly, even user intimate, and have the capability of allowing patrons to browse forward and backward through the authority files or the as-used headings." Horney,[52] in her 1985 advocation of "reasonable care" as opposed to perfection in creating automated bibliographic files, noted

In the early days of OCLC there was a widespread belief that computer access techniques would make consistency in heading control less important for an automated database than for the conventional card catalog. Experience has convinced most people that that idea was not well founded. Without some kind of cross-reference structure and the assurance of finding everything belonging under one heading with the same form of that heading, users of an online catalog are gravely misled. Finding some items under a particular heading, they are certain to think they have found everything appropriate.

Horney went on to conclude that "it is hard for libraries to take pride in computer databases full of misleading conflicts and uncollocated variant entries, since these are a clear disservice to the users who are the focus of our efforts." McDonald,[53] in a 1985 study of authority control and online catalogs, concluded

Authority control is not an anachronism of the precomputer age, but rather a tool employed throughout the data processing industry to manage large databases. As online catalogs proliferate, the importance and relevance of authority control increases in three ways. First, it is a tool that can be used to maintain the consistency and integrity of the online catalog. Second, it imposes a known structure on the online catalog, which facilitates information retrieval. And third, the combination of consistency, integrity, and structure enables systems to provide features that otherwise would not be possible.

Romero and Wajenberg,[54] in their 1985 article describing some of the problems involved in cataloging for the online catalog at the University of Illinois at Urbana-Champaign (UIUC), noted that "a major factor in selecting the WLN (Washington Library Network) programs for the UIUC online catalog was the provision they included for an authority file." Taylor, Maxwell, and Frost,[55] in their 1985 evaluation of authority control systems available from networks and vendors, stated that "network and vendor authority control is an idea whose time has come" and concluded that "automated authority control is emerging as one of the most fluid and exciting concepts in the building of total library systems." Tillett,[56] in her 1985 preliminary report of a survey examining the use of authority control among many different types and sizes of libraries, found that 13% of libraries surveyed had no authority files and noted that "the result is an incomplete catalog often lacking cross-references with maintenance of headings becoming a hit-or-miss activity since there is no record of which headings were used." Avram,[57] in her 1986 paper on the Linked Systems Project (LSP), noted that "the first application being developed for the LSP is the sharing of authority data." She also noted

The authority application was chosen as the first for LSP because the most expensive part of cataloging is creating headings and related cross-references for the building of consistent catalogs. In addition to the savings connected with the actual sharing of this data, the use of one consistent authority file by many institutions to establish headings for bibliographic records will make more effective over time the sharing of these bibliographic records without requiring local modification in order to add the records to the individual library catalogs.

Runkle,[58] in his 1986 advocacy of retrospective conversion of the Library of Congress catalog, asserted

Consistent entry forms and a system of cross-references have been the mechanical techniques traditionally used by librarians to create usable library catalogs. Even with catalogs in machine-readable format, authority control continues to be a requirement. I don't know of any evidence that demonstrates that computer technology obviates the need for authority control; I don't know of any large library that has plans to abandon authority control in its catalog.

Jamieson, Dolan, and Declerck,[59] in their 1986 examination of the need for authority control in the keyword searching environment of an online catalog, found that 53.7% of author headings, 73.5% of name subject headings, and 69.6% of topical subject headings could not be found by keyword searching and concluded "that a catalog without a cross-reference structure for variant forms of names and subject headings will give its users inferior service." Oddy,[60] in a 1986 address to the Cataloguing and Indexing Group of the Library Association, declared, "Any library which ignores authority control is seriously misusing and underusing its collection and is certainly not giving the users the service they deserve." Watson and Taylor,[61] in their 1987 study of the need for authority control in online catalogs, discovered that some 60% of the LC name authority file "would not have to be created in an online environment where keyword and automatic right-hand truncation searching capabilities are combined with computer file structures that do not require making an authority record for names not needing references" but still concluded, "Authority control continues to be the cornerstone of the catalog—manual or automated."

SURVEY OF LIBRARIES

A survey of American college and university libraries identified in *Library Statistics of Colleges and Universities, 1982 Institutional Data* as having library holdings in excess of 250,000 titles was conducted to ascertain their need for and use of authority control. Of the 301 questionnaires sent, a total of 171 libraries (57%) responded. The response was gratifying because most libraries responded fully and thoughtfully to the questions. The only exception was to the questions pertaining to the number of hours spent weekly to maintain authority control. Here, overall response was somewhat scanty with a considerable number of libraries admitting they did not keep track of this information.

The results were evaluated for overall response and then separated into categories related to online or manual catalogs and automated or manual authority control (see table 1). Most libraries (85%) had holdings of under two million titles. Thirty-seven percent possessed an online catalog; another 13% expected one shortly. Eighteen percent possessed some sort of automated authority control; an additional 12% had plans. Only 9% of the libraries claimed authority control linked to their online catalogs, but another 13% were in the planning stage. Approximately 15% of the libraries had closed catalogs for author, title, and subject upon adoption of AACR2. The percentage of libraries maintaining various forms of in-house authority control prior to AACR2 varied from a high of 87% for series titles to a low of 42% for uniform titles. There was a 10 to 16% increase in the number of libraries maintaining in-house authority control after adoption of AACR2, except for control of series and subjects, where little increase was evident. Over 90% of libraries utilize "see" references, while 80% utilize "see also" references. Ninety-one percent of all libraries utilized the OCLC bibliographic utility. Only 5% of the libraries anticipated abandoning authority control as a result of enhanced computer searching.

Online Catalogs with Linked Authority Control

There were fifteen libraries (9%) with online catalogs and linked authority control. As expected, only 3 to 11% of libraries with holdings within the 250,000 to 2,000,000 title range had these state-of-the-art catalogs, while 22 to 33% of larger libraries did possess them. Forty percent of libraries in this category had closed author, title, and subject catalogs upon adoption of AACR2. Libraries with online catalogs and linked authority control were more likely to have maintained higher levels of authority control for personal and corporate names prior to AACR2 and therefore showed no increase in these categories, but they did show an increase in authority control for conference names, geographic names, uniform titles, and subjects after the adoption of AACR2. Libraries in this category were less likely to utilize "see" references and more likely to include various forms of notes in their authority files. Of course those libraries possessing the current ulti-

Table 1. Survey of Libraries

	Overall (OA) (% of 171) No.	(%)	Online linked (% of OA) No.	(%)	Online unlinked (% of OA) No.	(%)	Manual (% of OA) No.	(%)
1. Libraries responding	171	(100)	15	(9)	48	(27)	108	(63)
2. Number of titles								
a. 250,000–500,000	65	(38)	3	(5)	8	(12)	54	(83)
b. 500,000–750,000	36	(21)	1	(3)	14	(39)	21	(58)
c. 750,000–1,000,000	17	(10)	1	(6)	5	(29)	11	(65)
d. 1,000,000–2,000,000	27	(16)	3	(11)	13	(48)	11	(41)
e. 2,000,000–3,000,000	12	(7)	4	(33)	5	(42)	3	(25)
f. 3,000,000–4,000,000	9	(5)	2	(22)	3	(33)	4	(44)
g. 4,000,000–5,000,000	1	—	0	(0)	0	(0)	1	(100)
h. Over 5,000,000	4	(2)	1	(25)	0	(0)	3	(75)
			(% of 15)		(% of 48)		(% of 108)	
3. Possess an online catalog	63	(37)	15	(100)	48	(100)	0	(0)
(in process)	23	(13)	0	(0)	0	(0)	23	(21)
4. Automated authority control	30	(18)	15	(100)	8	(17)	7	(6)
(will be)	21	(12)	0	(0)	8	(17)	13	(12)
5. Linked authority control	15	(9)	15	(100)	0	(0)	0	(0)
(will be)	23	(13)	0	(0)	10	(21)	13	(12)
6. Catalogs closed w/AACR2								
a. Author	26	(15)	6	(40)	13	(27)	7	(6)
b. Title	25	(15)	6	(40)	12	(25)	7	(6)
c. Subject	24	(14)	6	(40)	11	(23)	7	(6)
d. Shelf list	4	(2)	1	(7)	3	(6)	0	(0)
e. Name authority file	1	—	0	(0)	1	(2)	0	(0)
7. Authority control before AACR2								
a. Personal names	122	(71)	14	(93)	30	(63)	78	(72)
b. Corporate names	117	(68)	13	(87)	30	(63)	74	(69)
c. Conference names	94	(55)	10	(67)	26	(54)	58	(54)
d. Geographic names	86	(50)	11	(73)	26	(54)	49	(45)
e. Uniform titles	72	(42)	10	(67)	18	(38)	44	(41)
f. Series titles	149	(87)	13	(87)	41	(85)	95	(88)
g. Subjects	112	(65)	11	(73)	29	(60)	72	(67)
8. Authority control after AACR2								
a. Personal names	138	(81)	14	(93)	39	(81)	85	(79)
b. Corporate names	133	(78)	13	(87)	39	(81)	81	(75)
c. Conference names	117	(68)	13	(87)	37	(77)	67	(62)
d. Geographic names	112	(65)	13	(87)	36	(75)	63	(58)
e. Uniform titles	99	(58)	11	(73)	31	(65)	57	(53)
f. Series titles	153	(89)	13	(87)	41	(85)	99	(92)
g. Subjects	114	(67)	13	(87)	27	(56)	74	(69)
9. Include cross-references, etc.								
a. See (from)/Search under	159	(93)	13	(87)	44	(92)	102	(94)
b. See also/Search also under	137	(80)	12	(80)	35	(73)	90	(83)
c. History notes	86	(50)	10	(67)	26	(54)	50	(46)
d. Scope notes	66	(39)	11	(73)	18	(38)	37	(34)
e. Reference sources	56	(33)	9	(60)	17	(35)	30	(28)
f. Publisher	48	(28)	4	(27)	14	(29)	30	(28)
g. Call no.	75	(44)	6	(40)	22	(46)	47	(44)
h. Establishing authority	93	(54)	9	(60)	26	(54)	58	(54)
i. Catalogers' notes	106	(62)	10	(67)	36	(75)	60	(56)
10. Bibliographic utilities								
a. OCLC	156	(91)	14	(93)	43	(90)	99	(92)
b. RLIN	28	(16)	3	(20)	11	(23)	14	(13)
c. WLN	4	(2)	0	(0)	2	(4)	2	(2)
d. Other	4	(2)	1	(7)	2	(4)	1	(1)

Table 1. Cont.

e. More than one	19	(11)	2	(13)	9	(19)	7	(6)
11. Anticipate abandoning authority control due to enhanced computer searching								
a. Personal name	9	(5)	0	(0)	2	(4)	7	(6)
b. Corporate names	7	(4)	0	(0)	2	(4)	5	(5)
c. Conference name	8	(5)	0	(0)	2	(4)	6	(6)
d. Geographic names	7	(4)	0	(0)	2	(4)	5	(5)
e. Uniform titles	7	(4)	0	(0)	2	(4)	5	(5)
f. Series titles	12	(7)	1	(7)	2	(4)	9	(8)
g. Subjects	10	(6)	0	(0)	3	(6)	7	(6)
h. Maybe	6	(4)	0	(0)	2	(4)	4	(4)

mate automated authority control systems showed even less inclination toward abandoning authority control than did other libraries.

Online Catalogs without Linked Authority Control

There were forty-eight libraries (27%) with online catalogs but without linked authority control. This category included 29 to 48% of libraries with holdings of 500,000 to 4,000,000 titles. Although none of these libraries possessed authority control linked to their online catalogs, 21% were working toward that goal. Approximately 25% of the libraries in this group had closed author, title, and subject catalogs upon adoption of AACR2. These libraries showed the lowest level of authority control for personal and corporate names prior to AACR2 but after AACR2 showed an increase in authority control of from 18 to 27% in all categories except series and subjects. Libraries in this category were slightly less likely to use "see also" references and much more likely to include catalogers' notes in their authority files. These libraries were also more likely to subscribe to more than one bibliographic utility.

Manual Catalogs

There were 108 libraries (63%) with manual catalogs. Ninety percent of these libraries had holdings of less than two million titles, but this category also included most of the libraries (four out of five) with holdings in excess of four million titles. Twenty-three (21%) of these libraries are in the process of developing online catalogs, and thirteen (12%) are planning authority control linked to their online catalogs. Only 6% of these libraries had closed catalogs upon adoption of AACR2. This group showed the lowest increase in authority control with the advent of AACR2 but included the largest number of libraries (92%) concerned with maintaining authority control over series titles. These libraries were the least likely to subscribe to more than one bibliographic utility.

Additional Analysis

Comparison of limited and admittedly somewhat questionable data regarding hours spent weekly by various staff levels in the three libraries categories, without regard for size of library, shows an increase in time spent for authority control as the catalog/authority control system becomes more highly automated (see table 2). This finding was supported by a number of comments to the effect that maintaining authority control online was just as time-consuming as maintaining authority control manually. Hopefully, this suggests that automated authority control allows libraries to provide more extensive and effective authority control over bibliographic records, rather than implying that automation requires more time to maintain the same level of control. Justification for use of, as well as need for, authority control was overwhelmingly in support of the historical purposes (i.e., the finding and gathering functions of the library catalog). The need for collocation, consistency, cross-references, integrity, and uniformity in the catalog, along with improving user access, was repeated over and over again. No single group of libraries, except possibly the

Table 2. Hours Spent Weekly on Authority Control

Catalog/Authority control	Prof.	Para-prof.	Clerical	Total
Online/linked	32	63	51	146
Online/unlinked	30	37	21	88
Manual	14	17	27	58

very small or little used, seems immune from these needs regardless of level of automation. Some libraries noted that online catalogs are not fully effective without authority control. Perceived responsibilities toward cooperative automated systems was another reason given as justification for authority control. Justification for lack of authority control was extremely scanty and essentially linked to underfunding or the small size of a library. Only a few, mostly small libraries found the limited authority control inherent in the OCLC database adequate for their purposes.

URI AUTHOR/TITLE CATALOG

Since the catalog at the University of Rhode Island (URI) main library had been split in the mid-1970s, and entries for both subjects and series have been under full authority control (with maintenance of separate authority files) since that time, this study was restricted to a survey of corporate and personal names in the author/title catalog. LC cataloging records from the OCLC database are accepted by subprofessional staff without checking against any authority files. Incoming entries that clash during file revision with existing entries in the public catalog are checked against the OCLC online authority file and corrected. Entries in non-LC contributed records and original cataloging are checked against the OCLC online authority file by professional catalogers prior to acceptance/input on the OCLC system.

Cross-references for corporate and personal names have long been made on a highly selective and occasionally haphazard basis, but no separate name authority file has been maintained. "See" references are made for compound names, and prior to 1973 this reference was recorded in the public catalog as a tracing on the first card in a sequence of identical entries. "Most often used" AACR1 headings were revised to AACR2 form in 1980 by using the lists of name changes noted in the LC *Cataloging Service Bulletin*. Files of less than twenty entries were corrected to AACR2 form, and "see" references were occasionally made from the old form. Files of more than twenty entries were connected to the AACR2 form via "see also" references.

In spring 1987, shortly after this investigation was completed, the making of cross-references for personal names, corporate names, and uniform titles was greatly increased but still on a somewhat selective and haphazard basis (e.g., if a cross-reference files next to its referred entry, it is not made, and most foreign-language cross-references are not made), and a separate name authority file is not maintained. It was this knowledge of past actions, concern over current practice, and a curiosity as to how many discrepancies were being created in the URI author/title catalog that led the authors, in fall 1986, to conduct the following investigation.

Four hundred and fifty-four newly arrived OCLC cards requiring cross-references were intercepted prior to filing and compared to the online name authority file. Printouts containing cross-references for corporate and personal name entries on these cards were produced from the OCLC online authority file and then checked against the author/title catalog. The results were tabulated for the two categories of headings with regard to the presence of "see" or "see also" references and whether discrepancies (i.e., additional or split files) were being created (see table 3).

Analysis indicated that 42% of corporate name headings and 18% of personal name headings entering an online catalog, created from the existing manual catalog records (i.e., via archival tapes) without regard for authority control, could be expected to create additional legitimately related files totally unconnected by any syndetic structure. Lack of "see" references accounts for approximately 17% of the ad-

Table 3. Discrepancies between URI Author/Title Catalog and Newly Arrived OCLC Cards

	Name Headings					
Discrepancies	Corporate	(%)	Personal	(%)	Total	(%)
Number	131	(100)	323	(100)	454	(100)
"See" references[1]						
Potential	476		523		999	
Made correctly	8		18		26	
Made incorrectly	3		5*		8*	
Not made	465		503		968	
Additional files	23	(18)	55	(17)	78	(17)
"See also" references[2]						
Potential	32		4		36	
Made correctly	—		—		—	
Made incorrectly	—		—		—	
Not made	32		4		36	
Additional files	32	(24)	4	(1)	38	(8)
Additional files created	55	(42)	59	(18)	114	(25)
Additional files unrecognizable during manual revision	44	(34)	20	(6)	64	(14)

1. Always needed.
2. Needed if entries exist.
*Includes three references made outside the LC authority structure.

ditional files created for both corporate and personal names expected to enter an uncontrolled online catalog. Lack of "see also" references accounts for 24% of additional corporate name files but appears to be statistically insignificant in the creation of additional personal name files. Corrections initiated in a manual catalog at the time of filing do reduce the number of additional files created, but even so, 34% of corporate name headings and 6% of personal name headings entering the manual catalog are likely to create additional, legitimately related files not linked by cross-references. The average number of additional files created for both categories of headings was 25%, while the average number of additional unrecognizable files created was 14%. This compares favorably with the study by Dowell, which found that 14% to 20% of AACR2 headings entering an academic library catalog would conflict with existing headings.[62]

It would appear from a comparison with the survey of libraries (table 1) that the University of Rhode Island library is somewhat out of step with its peers. While most libraries with manual catalogs were and are more likely to maintain authority control for names rather than subjects, URI has long found it necessary to maintain full authority control over subjects. This reflects the needs of strong undergraduate programs in science and technology—a legacy of its land-grant original. With regard to authority control for series, URI is in step with its peers, the great majority of which seem to find series authority control more essential than authority control for any other category of entry.

CONCLUSIONS

Authority control is alive and well in most American college and university libraries. The survey of the literature produced a virtually unwavering sweil of support for the concept of authority control. The adoption of AACR2 caused ''braries to augment their previous efforts at authority control, while experience with online catalogs indicates the need for more, not less, authority control. The trend toward online catalogs with linked authority control is providing libraries with a greatly enhanced capability to control cataloging records and bodes well for patrons, who can look forward to better bibliographic access in the future than they have ever received in the past. And last, but certainly not least, those libraries, such as URI, accepting cataloging from the OCLC database without imposing full authority control over their records run a grave risk of losing control over their catalogs, be they manual or automated.

REFERENCES

1. Doris Hargrett Clack, "Authority Control: Issues and Answers," in *Libraries in the '80s*, ed. Dean H. Keller (New York: Haworth, 1985), p.127–40.
2. 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 (Mar. 1987).
3. Mitch Freedman, "A Conversation with Frederick G. Kilgour," *Technicalities* 1:1–20 (June 1981).
4. Audrey N. Grosch, "Computer-Based Subject Authority Files at the University of Minnesota Libraries," *Journal of Library Automation* 5:230–43 (Dec. 1972).
5. Frances Ohmes and J. F. Jones, "The Other Half of Cataloging," *Library Resources and Technical Services* 17:320–29 (Summer 1973).
6. S. Michael Malinconico, "The Role of a Machine Based Authority File in an Automated Bibliographic System," in *Foundations of Cataloging*, ed. Michael Carpenter and Elaine Svenonius (Littleton, Colo.: Libraries Unlimited, 1985), p.211–33. Published originally in *Automation in Libraries: Papers Presented at the CACUL Workshop on Library Automation, Winnipeg, June 22–23, 1974* (Ottawa: Canadian Library Assn., 1975).
7. S. Michael Malinconico, "The Library Catalog in a Computerized Environment," *Wilson Library Bulletin* 51:56–64 (Sept. 1976).
8. Hugh C. Atkinson, "The Electronic Catalog," in *The Nature and Future of the Catalog*, ed. Maurice J. Freedman and S. Michael Malinconico (Phoenix, Ariz.: Oryx, 1979), p.103–4.
9. Kenneth Bierman, "The Future of Catalogs in North American Libraries," in *The Nature and Future of the Catalog*, p.118.
10. Frederick G. Kilgour, "Design of Online Catalogs," in *The Nature and Future of the Catalog*, p.35, 38–39.
11. Anne Grodzins Lipow, "The Catalog in a Research Library and Alternatives to It," in *The Nature and Future of the Catalog*, p.246–47.
12. S. Michael Malinconico, "Bibliographic Data Base Organization and Authority File Control," *Wilson Library Bulletin* 54:36–45 (Sept. 1979).
13. Susan K. Martin, "Upgrading 'Brief and Dirty' Data," *American Libraries* 10:213–14 (Apr. 1979).
14. "Authority Control, or the Key to Survival in the Eighties," *Law Library Journal* 73:929–57 (Fall 1980).
15. Karen L. Matthews, "Authority Control," *Colorado Libraries* 6:16–20 (June 1980).
16. Jeanette Gail Mosey, "Name Authority Work in OCLC Libraries: A Survey of Practices and Expectations" (Ph.D. diss., Univ. of Southern California–Los Angeles, 1980), p.86–87.
17. Martin Runkle, "Authority in On-line Catalogs," *Illinois Libraries* 62:603–6 (Sept. 1980).
18. Helen F. Schmierer, "The Relationship of Authority Control to the Library Catalog," *Illinois Libraries* 62:599–603 (Sept. 1980).
19. Freedman, "A Conversation."
20. Ritvars Bregzis, "The Syndetic Structure of the Catalog," in *Authority Control: The Key to Tomorrow's Catalog*, ed. Mary Ghikas (Phoenix, Ariz.: Oryx, 1982), p.24, 26.
21. Grace Bulaong, "Authorities and Standards in a Changing World," *International Cataloguing* 11:35–36, 41–44 (July/Sept.–Oct./Dec. 1982).
22. James D. Craig, "Series Authority Files: The Glasgow University Experience," *Journal of Librarianship* 14:289–96 (Oct. 1982).
23. Gwen Miles Culp, "Authority Control within the Washington Library Network Computer System," in *Authority Control: The Key to Tomorrow's Catalog*, p.72.
24. Arlene Taylor Dowell, *AACR 2 Headings: A Five-Year Projection of Their Impact on Catalogs* (Littleton, Colo.: Libraries Unlimited, 1982), p.104.
25. Michael Gorman, "Authority Control in the Prospective Catalog," in *Authority Control: The Key to Tomorrow's Catalog*, p.170–71.
26. Lorene E. Ludy and Susan J. Logan, "Integrating Authority Control in an Online Catalog," in *Proceedings of the ASIS Annual Meeting* 19:176–78 (1982).
27. Mary A. Madden, "Is This Somehow Connected? The Vendor Perspective," in *Authority Control: The Key to Tomorrow's Catalog*, p.94.
28. Susan Matson, "Desiderata for a National Series Authority File," *Library Resources and Technical Services* 26:331–44 (Oct.–Dec. 1982).

29. R. Bruce Miller, "Authority Control in the Network Environment," in *Authority Control: The Key to Tomorrow's Catalog*, p.42, 45.
30. Jean M. Perreault, "Authority Control, Old and New," *Libri* 32:124–48 (June 1982).
31. Lucia J. Rather, "Authority Systems at the Library of Congress," in *Authority Control: The Key to Tomorrow's Catalog*, p.163.
32. Nancy J. Williamson, "Is There a Catalog in Your Future? Access to Information in the Year 2006," *Library Resources and Technical Services* 26:122–35 (Apr.–June 1982).
33. Carolyn W. Baker, "Name Authority and Arkansas Academic Libraries," *Arkansas Libraries* 40:8–9 (Mar. 1983).
34. Cathy Ann Elias and C. James Fair, "Name Authority Control in a Communication System," *Special Libraries* 74:289–96 (July 1983).
35. Ruth Patterson Funabiki, Ingrid Mifflin, and Karen Corlee, "Use of the WLN Authority Control System by an ARL Library," *Library Resources and Technical Services* 27:391–94 (Oct.–Dec. 1983).
36. Kathryn Luther Henderson, "Great Expectations: The Authority Control Connection," *Illinois Libraries* 65:334–36 (May 1983).
37. Arnold S. Wajenberg, "The Use of LC Authority Records in the On-line Catalogue of the University of Illinois at Urbana-Champaign," *Illinois Libraries* 65:331–33 (May 1983).
38. Henriette D. Avram, "Authority Control and its Place," *Journal of Academic Librarianship* 9:331–35 (Jan. 1984).
39. Robert H. Burger, "Artificial Intelligence and Authority Control," *Library Resources and Technical Services* 28:337–45 (Oct./Dec. 1984).
40. Jean Dickson, "An Analysis of User Errors in Searching an Online Catalog," *Cataloging & Classification Quarterly* 4:19–38 (Spring 1984).
41. Katharina Klemperer, "Authority Control in the University of California MELVYL System," in *Crossroads: Proceedings of the First National Conference of the Library and Information Technology Association*, September 1983, ed. Michael Gorman (Chicago: American Library Assn., 1984), p.124.
42. Suzanne L. Liggett, "The Name Authority Co-op Project at the Library of Congress," in *Crossroads: Proceedings of the First National Conference of the Library and Information Technology Association*, September 1983, ed. Michael Gorman (Chicago: American Library Assn., 1984), p.121.
43. Lorene E. Ludy and Sally A. Rogers, "Authority Control in the Online Environment," *Information Technology and Libraries* 3:262–66 (Sept. 1984).
44. Dan Miller, "Authority Control in the Retrospective Conversion Process," *Information Technology and Libraries* 3:286–92 (Sept. 1984).
45. Erlene Rickerson, "The Washington Library Network Automated Authority Control System," in *Crossroads: Proceedings of the First National Conference of the Library and Information Technology Association*, September 1983, ed. Michael Gorman (Chicago: American Library Assn., 1984), p.133.
46. Melinda L. Shore, "Variation between Personal Name Headings and Title Page Usage," *Cataloging & Classification Quarterly* 4:1–11 (Summer 1984).
47. Arlene G. Taylor, "Authority Files in Online Catalogs: An Investigation of Their Value," *Cataloging & Classification Quarterly* 4:1–17 (Spring 1984).
48. Catherine M. Thomas, "Authority Control in Manual Versus Online Catalogs: An Examination of 'See' References," *Information Technology and Libraries* 3:393–98 (Dec. 1984).
49. Robert H. Burger, *Authority Work: The Creation, Use, Maintenance, and Evaluation of Authority Records and Files* (Littleton, Colo.: Libraries Unlimited, 1985), p.10.
50. Clack, "Authority Control," p.127, 139.
51. Susan Baerg Epstein, "Automated Authority Control: A Hidden Timebomb?" *Library Journal* 110:36–37 (Nov. 1, 1985); 111:55–56 (Jan. 1986).
52. Karen L. Horney, "Quality Work, Quality Control in Technical Services," *Journal of Academic Librarianship* 11:206–10 (Sept. 1985).
53. David R. McDonald, "Data Dictionaries, Authority Control, and Online Catalogs: A New Perspective," *Journal of Academic Librarianship* 11:219–22 (Sept. 1985).
54. Nancy Romero and Arnold Wajenberg, "Authority Records and Authority Work in the Online Catalogue," *Information Technology and Libraries* 4:318–23 (Dec. 1985).
55. Arlene G. Taylor, Margaret F. Maxwell, and Carolyn O. Frost, "Network and Vendor Authority Systems," *Library Resources and Technical Services* 29:195–205 (Apr.–June 1985).
56. Barbara B. Tillett, "1984 Automated Authority Control Opinion Poll: A Preliminary Analysis," *Information Technology and Libraries* 4:171–78 (June 1985).
57. Henriette D. Avram, "The Linked Systems Project: Its Implications for Resource Shar-

ing," *Library Resources and Technical Services* 30:36–46 (Jan.–Mar. 1986).

58. Martin Runkle, "Authority Control: A Library Director's View," *Journal of Academic Librarianship* 12:145–46 (July 1986).
59. Alexis J. Jamieson, Elizabeth Dolan, and Luc Declerck, "Keyword Searching versus Authority Control in an Online Catalog," *Journal of Academic Librarianship* 12:277–83 (Nov. 1986).
60. Pat Oddy, "Name Authority Files," *Catalogue & Index* 82:1,3–4 (Autumn 1986).
61. Watson and Taylor, "Implications of Current Reference, p.18."
62. Dowell, "AACR 2 Headings." ■■

Twenty Years Ago in JOLA

While the number of articles on catalog conversion has not yet been overwhelming, it is apparent that there is a great deal of interest in the field. One might ask: "Should every library proceed to convert its own catalog?" . . . Perhaps the question has no ideal answer. It does seem unfortunate, however, that the pre-1955 National Union Catalog is not to be published from a record that is machine readable.

It would seem possible, however, that in the future methods could be devised to use machine readable records produced by the larger libraries and some procedure whereby the smaller libraries could check their holdings against those of the larger libraries. By some fairly simple method, a subset of the master machine records could be selected for use in the catalogs of smaller libraries.

Donald V. Black, "Creation of Computer Input in an Expanded Character Set," *JOLA* 1, no.2:119 (June 1968).

Management Information for the CIP Program Using SAS

Ruta Pempe Godwin and Susan H. Vita

A management information system gives continual feedback on work flow and changes in operations. Management reports make it possible to evaluate unintended trends as well as planned changes and to plan for future operations with good historical data.

Management statistics are one of the benefits we expect to realize from putting records into machine-readable form. With such powerful software tools as the Statistical Analysis System (SAS) package, the effort (and cost) of developing software to produce specific reports is far less now than when totally customized programs were the only alternative.

This article is about such an MIS system—how management information needed by the CIP program was analyzed and a system constructed to produce information about the work flow in a cataloging program. Certain aspects of this project may be of general interest. The source of the statistical data is a bibliographic record in the MARC format. The techniques used to arrive at requirements and specifications, the technical strategy, and the programming effort are described. The resulting system, which is fairly easy to implement and easy to use, incorporated the powerful capabilities of a widely available statistical analysis software product.

CIP AND ITS SERVICES

The Cataloging in Publication (CIP) Division of the Library of Congress administers the CIP program, a cooperative effort between the library and U.S. publishers for the purpose of preparing prepublication catalog records for forthcoming monographic titles.[1]

Publishers participating in this program submit manuscripts, galleys, or front matter of a forthcoming title to the CIP division. Based on this prepublication information, the library's professional cataloging staff creates a bibliographic record. Within ten working days after receipt of information from the publisher, the CIP division sends back an abbreviated version of this record to be printed in the book. Upon publication, a copy of the book must be sent to the CIP division. The library uses these books to correct the cataloging data, if necessary, then processes the books for the library's collections. All monographic trade publications published in the U.S. (including textbooks above the high-school level) and monographic federal government publications are within the scope of CIP.

The CIP division currently has about thirty staff members. They use several microcomputer workstations for general office automation and about ten terminals to

---

Ruta Pempe Godwin is formerly Senior Automation Planning Specialist, Automation Planning and Liaison Office, Library of Congress. She is currently Project Leader, Office of Information Systems, National Science Foundation. Susan Vita, formerly Chief, Cataloging in Publication Division, is currently Chief, MARC Editorial Division, Library of Congress.

access the library's systems.

CIP cataloging is created in the library's cataloging divisions, and the final CIP record is input to the library's automated cataloging system by the MARC editorial division. Active CIP records are included in the online catalog—accessible everywhere within the Library of Congress—and distributed to libraries nationwide through the MARC Cataloging Distribution Service. Libraries use the information to order new titles and to process their new publications; therefore, the more titles in the CIP program, the better it is for the library community.

NEED FOR MANAGEMENT INFORMATION

The current volume of CIP titles (i.e., 40,000 per year) requires management through several processing cycles, some of them time critical, to meet the program agreements and needs of the Library of Congress collections. The program has grown rapidly in the last ten years, and management information has been essential for continual feedback on work flow in order to plan changes in procedures and staffing and to help ensure that the program continues to meet the needs of those using CIP records. CIP managers felt that the program's size, although welcomed as a sign of its success, hindered their ability to manage this information efficiently in a manual environment. Therefore, it was decided that an automated management information system was needed.

Management information about the program must include statistics on the volume of submissions received from each publisher, the number of publishers participating in the program, the number of titles received for LC collections, the number of claims issued for books that are published but have not been sent to LC by the publishers, the timing between a prepublication CIP submission and the receipt of the published work, the number of pre- and postpublication changes (and their types) made to CIP cataloging records, and so on.

Each of these statistics has multiple uses—in determining work loads and allocating staff resources and in monitoring and evaluating the CIP program's performance in fulfilling its mission to the library community. For example, the total number of titles submitted by each publisher is compared with the publisher's output to determine whether publishers are generally submitting all eligible titles for the CIP program. Those who are not can be contacted and urged to increase their participation. The numbers and types of postpublication changes are tracked to help determine staffing work loads, whether publishers need reminders of their obligation to inform the CIP division of changes made prior to publication so that information can be distributed to the library community, and whether each publisher is submitting enough information to make cataloging decisions at the prepublication stage. Information on timing between CIP submissions and receipt of the work and claiming information is needed to determine staffing work loads and patterns in publishers' conformance with the obligation to send LC a copy of the work so that the cataloging record can be verified and completed.

A management information system that would allow the CIP division to monitor its operations closely and help in the continual reevaluation of the effectiveness of the program was determined to be essential.

OUR APPROACH TO REQUIREMENTS

Requirements were developed by CIP managers and a systems analyst. First, the managers held brainstorming sessions to identify statistics—either currently unavailable or only available through laborious manual routines—that would help them better manage the CIP program and its operations. About forty questions were identified, e.g., What is the average number of CIP records per publisher per year? How many publishers actively participate in the CIP program in a given year? How many CIP records are updated due to a change in the projected publication date (communicated from publishers)? What is the breakdown by Dewey classification of all CIP records processed in the last year?

The systems analyst created a fact sheet on every target report as a way of gathering information for specifications, a "bottom up" approach (see figure 1). The team discussed how the statistics might be derived. For example, the number of publishers participating in the CIP program could be known by counting the number of entries in a manual name and address file, the number could be derived via a query to a planned microcomputer publisher directory system, or it could be derived by running a mainframe program to count the number of discrete publishers represented in a machine-readable file of all CIP records processed in a given year.

The fact sheet summarized various aspects of each target statistic: What is the management question answered by the statistic? What are sources of raw data for the statistic? What is the logical derivation of the statistic from raw data? What should the report look like (the title and format of the report)? and What could a graphic representation look like? The fact sheets made it possible to group target statistics by type of raw data and by type of management information.

This method allowed us to discuss each report as a specific component of management information. We could model the report and define it in terms of instructions to a programmer. Reports that had to be produced on the same schedule, depended on the same machine-readable data, or required similar programming logic could be bundled for implementation. For example, some reports required a full year of new CIP records in order to derive a complete analysis by subject.

We also found that reports pertaining to publishers would require linking the imprint information contained in CIP MARC records—the publisher name and address—with additional information.[2] An electronic file containing information about publishers was planned to build the potential to analyze. For example, What is the geographic distribution of publishers participating in the CIP program? How many publishers are in the program (using a definition of *publisher* according to the CIP program rather than counting the unique names used in cataloging)? and so on.

We expected that many of the target statistics could be derived from machine-readable CIP records. The machine-readable CIP record (online) contained information—e.g., date entered file, date of last transaction, record status (new or updated), publisher name, Library of Congress Control Number (LCCN)—that

CIP Statistics Fact Sheet

Code Name:	CC04
Group:	CIP claiming
Title:	Numbers of CIP records for which items are received
Description:	How many items are received via the CIP program, to date and by month?
Source system:	SAS cumulated statistics file
Source record:	CIP statistical record
Schedule:	On demand
Target statistics:	Number of CIP records that have PPD='1111' (received item), by month in which received
Derivation spec:	Select all records that are changed CIPs (RECTYPE='10') and have a PPD='1111'
	Dedupe them based on LCCN (want to count only one per LCCN)
	Count records, report the total Count by month, based on Date of Last Transaction (DLT)
Process in doc:	Process 8
Example of graph:	Attached

Fig. 1. Fact Sheet for MIS Requirements.

could be used to count records within a certain date range, numbers of records per publisher, records with certain Dewey numbers, and so on.

Once we discovered that much information was potentially available from the online system used to process CIP cataloging, the next question was, How to derive summary statistics?

Many of the statistics we sought could be acquired by recording and counting data from each new CIP record created within the last year: its LCCN, publisher, projected publication date, Dewey number, and so on. Other statistics, however, could only be derived by comparing the original CIP record with its later update(s). For example, all statistics relating to CIP record changes depended on comparing the original record with later versions that might show that the publication date has been changed or that the published CIP book had been received.

IMPLEMENTATION OPTIONS

One option was to request the enhancement of the online system so that every time a CIP record was input or updated a program would monitor the transaction and record the information for statistical purposes. This option was not practical for a number of reasons: most importantly the task did not have a high priority and thus would not be implemented in the near future.

Another option was to use batch programs to analyze CIP records rather than analyzing them at each transaction. The back-up files (on tape) of the entire online database contain CIP MARC records along with records for other cataloging. These tapes can be scanned for CIP records. A hazard of this approach is that the "electronic review" in batch mode would probably take place less frequently than once a day. However, managers knew from experience that a CIP record (or other LC cataloging record) was not likely to be updated more than once per day. A CIP record is normally upgraded fully when the title is received in cataloging divisions. The full processing cycle, which depends on publishing schedules, takes months, not days. Consequently, we decided a monthly "review" was enough to capture most updates.

For our purposes we decided that a brief, fixed-length record containing the raw data essential for all our target reports could be more easily analyzed than a full MARC record (see figure 3). A monthly scan of all CIP records would extract information from all records that were added or updated in the last month. The file of extracted information would cumulate throughout the year, strictly for statistical purposes.

The Library of Congress has a powerful retrieval program that can identify, copy, and count our internal MARC-like machine-readable bibliographic records based on nearly any MARC data content or structural aspect. The program, JANUS, uses an end-user oriented language that specifies exactly which fields are to be examined (e.g., title, author, LCCN, subject fields, local fields) or which structural attributes (e.g., length, number of fields, subfields) a record should have.[3] JANUS queries are contained in parameter cards in a batch program.

Rather than write a special program to select CIP records from the cataloging file, we were able to use JANUS and supply it with a one-line query. The query basically instructed JANUS to select CIP records with activity in the last month by asking for records that "have tag-263 or tag-905 and date-of-last-transaction greater-than-today-minus-30-days," for example:

```
MATCH (TAG = 263 ! = 905) AND
DLTP > = TODAY–30;
```
[4]

Once a month, the JANUS program is used to select a copy of all CIP records with recent transactions. Thus, using JANUS we could maintain a file containing copies of every CIP record in its original MARC-like format and all its subsequent updates over the year and count these records.

JANUS, however, had two limitations that prevented it from being the only system we needed to achieve our ends. First, it could not perform statistical calculations on the raw data. For example, to know how long in advance of publication a CIP record had been requested, we needed to subtract the Date Entered Masterfile from the Projected Publication Date (in a YYMM format) and use that "Time-In-Advance" for an analysis of "Time-In-Advance" across all CIP records. JANUS could not make these arithmetic calculations. Second, JANUS currently has no graphic capability, so in cases where it could count—for example, all CIP records in Dewey number ranges—it could not present the numbers as a graph. We could have obtained the graphic presentation, of course, by rekeying the summary statistics into a microcomputer package such as a spreadsheet or a charting program, but this option seemed too inefficient.

Our solution was to maintain a file of short, fixed-length records that represent each new version of a CIP record and contain the data needed for statistical analysis, using a generalized software package—SAS—to analyze the records.[5]

DEVELOPING THE SOFTWARE

The software we needed could be divided into four components:

1. An online interface—a gateway—that would allow a CIP staff member to request certain reports on demand. The interface on our systems would use RPF, or ROSCOE Programming Facility, which is a very high-level language for writing macros under ROSCOE, the telecommunications facility on our mainframe computers.

RPF makes it possible to present menu screens and to prompt for operator selections and command words, if necessary, then initiate a batch job submission (see figure 2).

2. A conversion program that would take the full MARC-like record in LC internal format and produce from it a fixed-length record (see figure 3).

The conversion program had to be written in PL/1 so that it could be maintained by Library of Congress programmers.

3. SAS programs that would read the fixed-length "shadow record" and analyze the raw data to derive target statistical reports.

4. Files to hold the cumulated statistics and their backups cataloged on the computer system.

The entire programming, which took 200 hours of staff time was done by experienced programmers and familiar with the particular software tools used.[6]

SAMPLE REPORTS

Figures 4 through 12 show examples of some of the reports that are part of the CIP management information system, and the examples illustrate some of the design principles we followed. For example, many reports break down categories by month. No matter when the program is run, the current month's profile and all previous months in the current year will appear in the analysis. This avoids a dependence on scheduling the computer batch job.

CIP managers must, of course, interpret the tables and charts output by the system. The reports aid in monitoring normal patterns by presenting data in graphic form so

```
                    CIP STATISTICAL MENU

   USING THE TAB KEY, PLACE THE CURSOR IN FRONT OF THE DESIRED
   REPORT AND PRESS THE ENTER KEY.

   ADD THE LAST MONTH OF CIP DATA TO STATISTICS FILE
   ARCHIVE ANNUAL CUM FILE AND START NEW CUM FILE
   STATISTICAL REPORT OF RECORD BY TYPE
   STATISTICAL REPORT ON CLAIMING
   STATISTICAL REPORT ON RECORDS PROCESSING
   STATISTICAL REPORT ON PUBLISHER ACTIVITY

   EXIT
```

Fig. 2. Gateway to MIS Reports.

MARC ELEMENT	NAME	FIXED LENGTH
001	LCCN	8
260 $b	Imprint, Publisher Name	40
263 $a	Projected Publication Date	4
020 $a	ISBN	10
050 $a	LC Call Number	9
060 $a	NLM Call Number	15
082 $a	Dewey Decimal Classification	10
086 $a	Sudoc Number	18
008,20	Intellectual Level Code	1
008,5-8	Date 1 (Primary Date)	4
RCA,0	Record Status Flag	1
RCA,1	Record Type Flag	1
RCA,2-3	Date of Last Verification	6
RCA,4-5	Date Entered Masterfile	6
RCA,6-7	Date of Last Transaction	6
	Total length	160 bytes

Fig. 3. Sample of Data Elements Copied from the MARC Record.

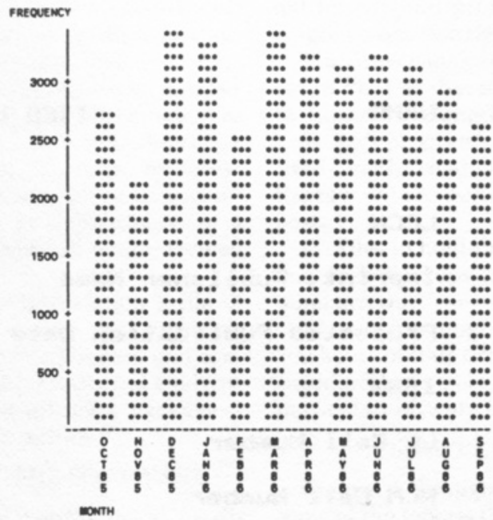

Fig. 4. Number of New CIP Records Input by Month.

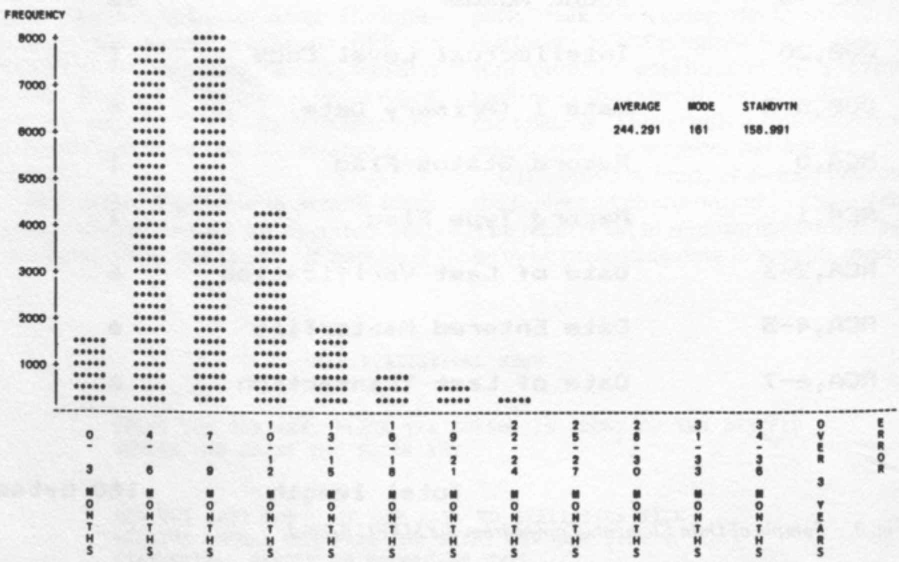

Fig. 5. Average Months from Time CIP Record Entered File to Receipt of Book.

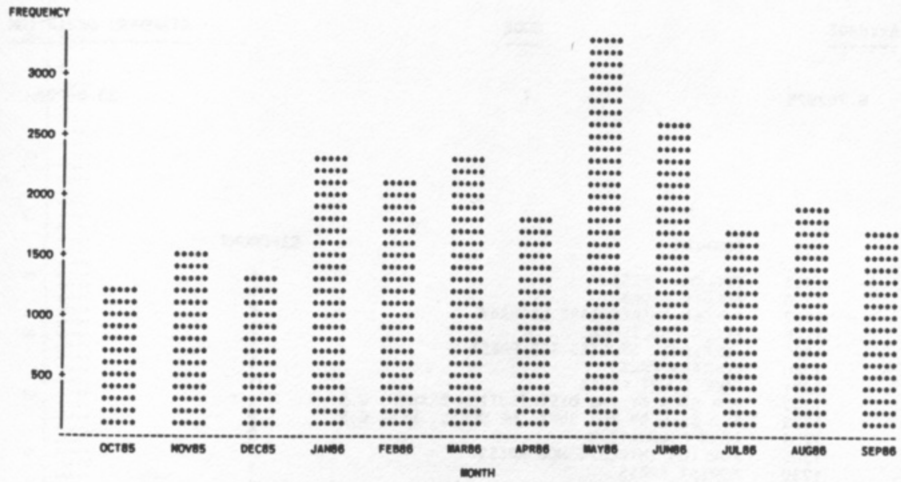

Fig. 6. CIP Records Received by Month Received.

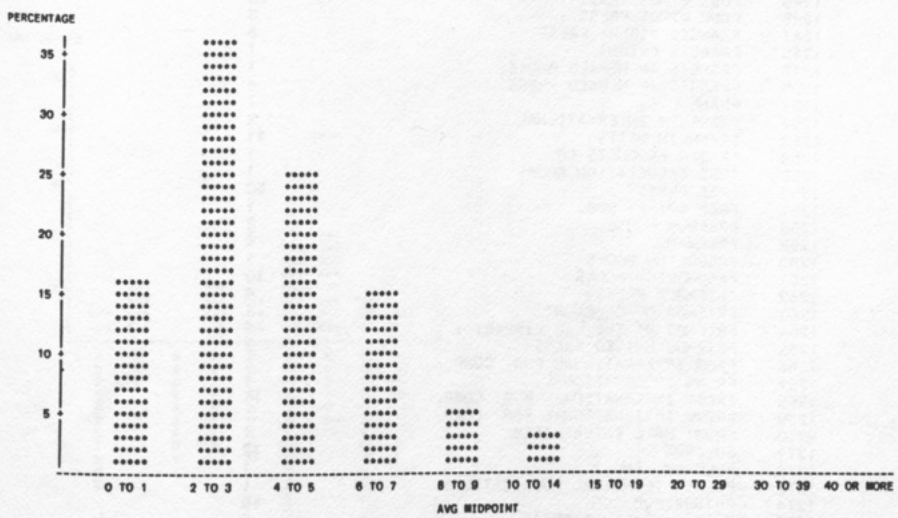

Fig. 7. Number of Months in Advance of Proposed Publication Date Titles Are Sent to CIP Program.

AVERAGE	MODE	STANDARD DEVIATION
9.262925	1	33.94792

OBS	PUBNAME	CIPCOUNT
1225	FOGHORN PRESS	1
1226	FOLGER BOOKS	1
1227	FOLGER SHAKESPEARE LIBRARY	4
1228	FOLIO PUBLISHERS	1
1229	FOLKLORIC STUDIES TGB PRESS	1
1230	FONTANA PRESS	1
1231	FOOL COURT PRESS	1
1232	FOR SALE BY THE DISTRIBUTION BRANCH, U.S	1
1233	FOR SALE BY THE SUPT. OF DOCS., U.S. G.P	3
1234	FORD FOUNDATION	2
1235	FOREIGN INTELLIGENCE PRESS	1
1236	FOREST PRESS	1
1237	FOREVERLY MUSIC	1
1238	FORTRESS PRESS	73
1239	FORUM PRESS	3
1240	FOUNDATION CENTER	1
1241	FOUNDATION FOR LIFE ACTION	5
1242	FOUNDATION PRESS	28
1243	FOUNDERS SOCIETY, DETROIT INSTITUTE OF A	1
1244	FOUR QUARTERS PUB. CO.	1
1245	FOUR WINDS PRESS	23
1246	FOUR WINDS PRESS :	3
1247	FRANCIS ASBURY PRESS	3
1248	FRANCIS CAIRNS	1
1249	FRANCISCAN HERALD BOOKS	1
1250	FRANCISCAN HERALD PRESS	4
1251	FRANK CASS	1
1252	FRANKLIN INTERNATIONAL	2
1253	FRANKLIN WATTS	11
1254	FRASER PRODUCTS CO.	1
1255	FREE ASSOCIATION BOOKS	1
1256	FREE PRESS	50
1257	FREE SPIRIT PUB.	1
1258	FREEDOM HOUSE	3
1259	FREEMAN	5
1260	FREUDLICH BOOKS	1
1261	FREUNDLICH BOOKS	10
1262	FRIENDLY PRESS	4
1263	FRIENDS OF CLERMONT	1
1264	FRIENDS OF THE LSU LIBRARY :	1
1265	FRIENDS UNITED PRESS	2
1266	FROM INTERNATIONAL PUB. CORP.	1
1267	FROMM INTERNATIONAL	1
1268	FROMM INTERNATIONAL PUB. CORP.	15
1269	FROMM INTERNATIONAL PUB. CORP. :	1
1270	FRONT HALL ENTERPRISES	1
1271	FULCRUM	12
1272	FULCRUM, INC.	1
1273	FULLER TECHNICAL PUBLICATIONS	1
1274	FUTURA PUB. CO.	16
1275	G¯IT¯A-N¯AGAR¯I PRESS	2

Fig. 8. Number of CIP Records per Publisher.

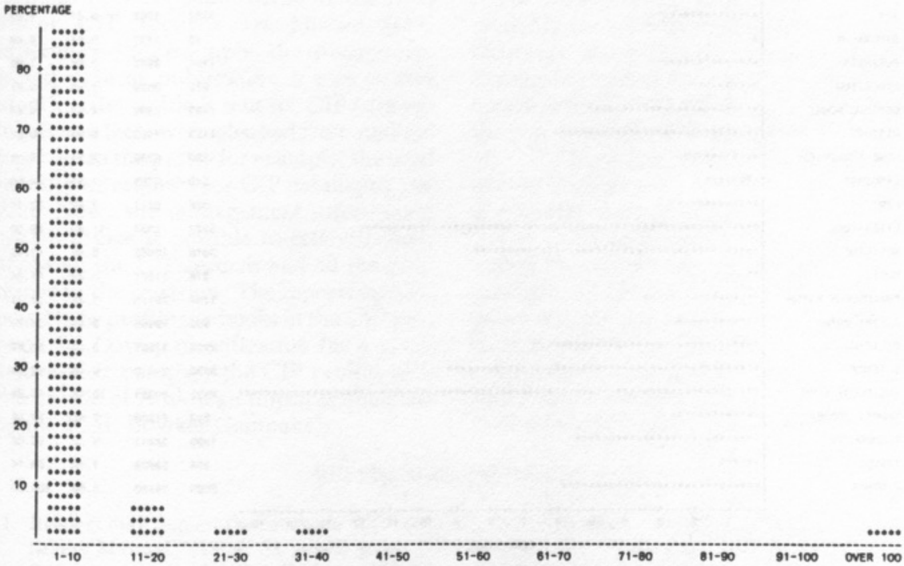

Fig. 9. Number of CIP Records per Publisher.

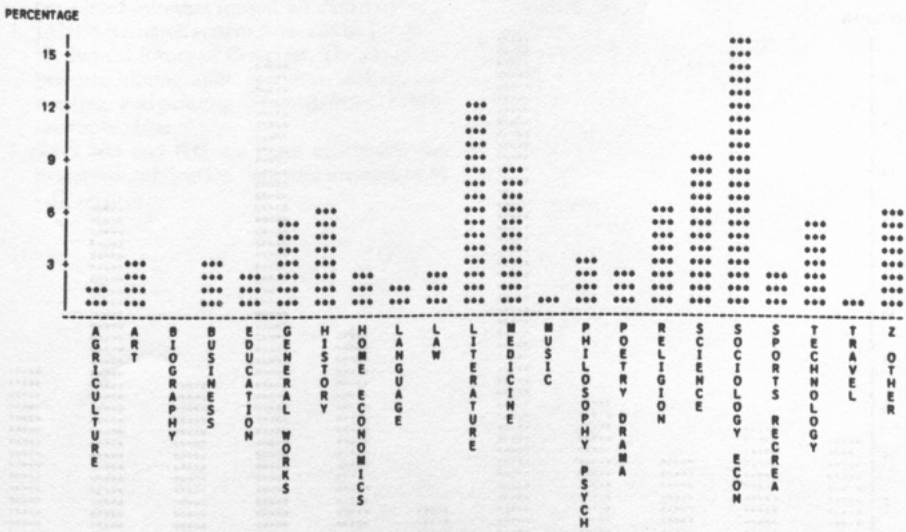

Fig. 10. CIP Records by Bowker Dewey Categories.

RANGEB	FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
AGRICULTURE	480	480	1.34	1.34
ART	1182	1662	3.31	4.65
BIOGRAPHY	76	1738	0.21	4.86
BUSINESS	1104	2842	3.09	7.95
EDUCATION	678	3520	1.90	9.85
GENERAL WORKS	1782	5302	4.99	14.84
HISTORY	2153	7455	6.03	20.86
HOME ECONOMICS	790	8245	2.21	23.06
LANGUAGE	540	8785	1.51	24.59
LAW	926	9711	2.59	27.18
LITERATURE	4223	13934	11.82	39.00
MEDICINE	3019	16953	8.45	47.45
MUSIC	318	17271	0.89	48.34
PHILOSOPHY PSYCH	1204	18475	3.37	51.71
POETRY DRAMA	890	19365	2.49	54.20
RELIGION	2222	21587	6.22	60.42
SCIENCE	3253	24840	9.10	69.52
SOCIOLOGY ECON	5523	30363	15.46	84.98
SPORTS RECREA	923	31286	2.58	87.56
TECHNOLOGY	1955	33241	5.47	93.03
TRAVEL	394	33635	1.10	94.14
Z OTHER	2095	35730	5.86	100.00

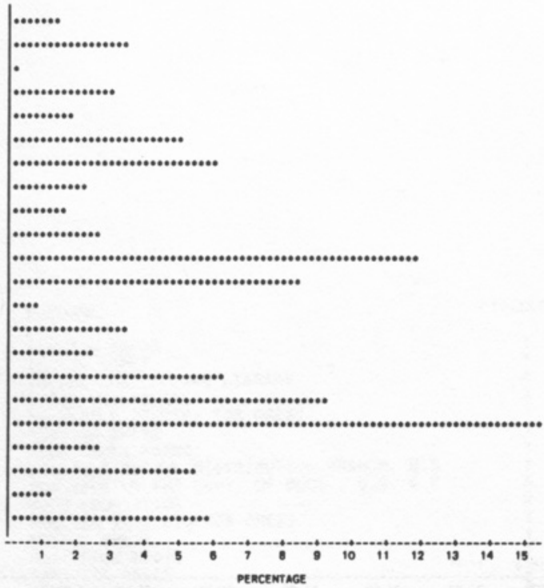

Fig. 11. CIP Records by Bowker Dewey Categories.

Figures 10–11 may be compared with Bowker's yearly book output statistics to determine how well the CIP program is doing in including the annual output of U.S. publishers.

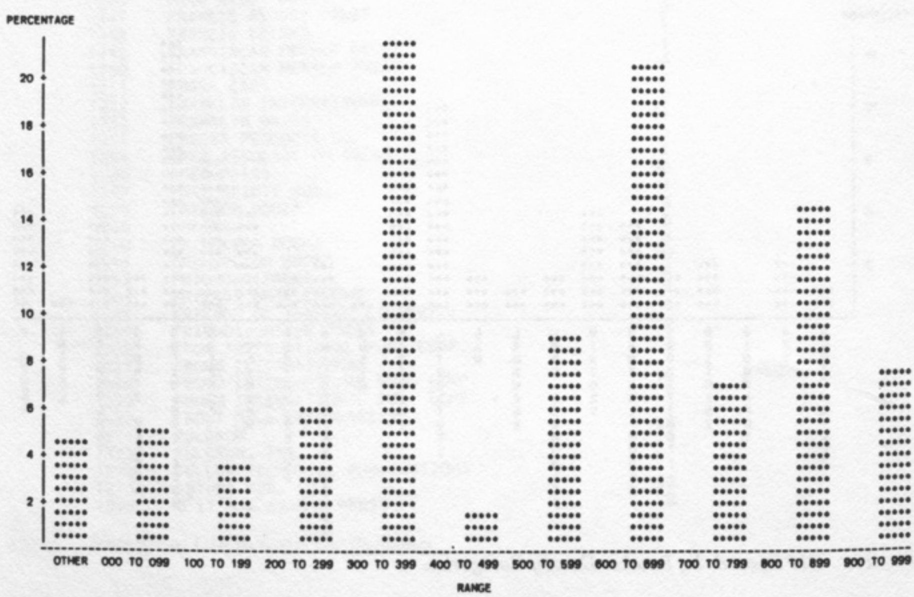

Fig. 12. CIP Records by Dewey Number Blocks (Hundreds).

that changes from the normal pattern of work are clearly visible. A change may mean that the CIP division should expect changes in work load. Some of the facts about CIP records are known from experience—for example, the average projected date of publication is two to five months after a title is sent for CIP cataloging. Some facts were gleaned from manual logs kept in the past—for example, the total number of requests for CIP cataloging per year. Now, the management information system makes it possible to establish facts about all the CIP records and all the publishers in the program. The reports make it possible to profile the books in the CIP program by Dewey classification for a given year and to compare the CIP profile with the profile of American imprints analyzed by the R. R. Bowker Company.

CONCLUSIONS

Libraries with access to SAS and with copies of their library catalogs in machine-readable form could implement a similar technique to produce management reports that are not part of their online systems. Libraries with OCLC-MARC tapes are using the tapes for collection analysis, for example.[7] We have described using online records to analyze work flow and the content of a special cataloging program.

The CIP program now has a software system on which to build other statistical products, as needs are justified. CIP program managers will benefit from having facts about the program to balance and complement managerial vision, experience, and intuition—the old-fashioned "management information systems."

REFERENCES AND NOTES

1. Readers may request the brochure "CIP Publishers Manual" (1984; 38 pages) from the Cataloging in Publication Division, Library of Congress, Washington, DC 20540.
2. The CIP record is initially in an LC internal data-storage format similar to the U.S. MARC communications format and may be converted into that format for distribution.
3. JANUS is a batch system authored by James L. Godwin, Library of Congress. The programs perform bibliographic retrieval, sorting, formatting, and printing of records from LC bibliographic files.
4. Tags 263 and 905 are fields containing the projected publication data and are unique to CIP records.
5. For a description of the software, see the *SAS Users Guide: Basics* 1984 or later edition (Cary, N.C.: SAS Institute, 1984).
6. A similar project contracted out to a data-processing vendor will very likely take longer. The effort depends on how familiar programmers are with MARC-formatted data, the programming tools, and the system environment in which the programs will run.
7. See Evelyn Payson and Barbara Moore, "Statistical Collection Management Analysis of OCLC-MARC Tape Records," *Information Technology and Libraries* 4:220–32 (Sept. 1985).

Expert Systems for Automatic Cataloging Based on AACR2: A Survey of Research

Roy Meador III and Glenn R. Wittig

The research literature on the subject of expert system use for descriptive cataloging is reviewed. Reasons why automatic cataloging is not as easy to establish as it would seem are explored, and implications of these investigations in regard to current cataloging rules are suggested.

Research into expert systems can be traced back to Alan Turing and others who explored the general problem of designing "intelligent" machines. Many of the first artificial intelligence (AI) programs were game-playing systems; they were inflexible and hardly resembled the knowledge of an expert. A computer that could "play" chess, for instance, even if it had the ability to "learn," was incapable of matching the ability of a chess master. The master's intuitive grasp of the whole game, his ability to focus on only the best alternatives at each point in the game, his ability to recognize patterns—all of which could be done without having to review each rule and each game ever played—seemed impossible to build into the linear, rule-based system of a computer.

Today expert systems employing sophisticated and complex heuristic capabilities easily approximate the knowledge of experts. There are now programs ranked on the expert level for chess and backgammon. But as AI research continues, machines are no longer just playing games: medical, business, and scientific applications are quite common and are multiplying at a surprising rate. In librarianship, the majority of the research and all of the applications pertain to online searching. Other areas of application in our field have been ignored or left untouched. This situation is disconcerting, particularly in cataloging, where a code (AACR2) is available and where algorithms for some routines are already in place.

With only a cursory look at the literature, it is evident that the term "expert system" is used loosely and, consequently, ambiguously. Moreover, since research into the use of expert systems for descriptive cataloging is in its early stages, little about it has been published. The main problem is piecing together this ongoing work to get an overall picture of automatic cataloging based on AACR2. The purpose of this article is to explain briefly the principles of expert systems and their uses and then to survey some of the research that has been completed or is still in progress.

EXPERT SYSTEMS

Expert systems are knowledge-based computer programs capable of solving complex problems that only an expert in a particular field could solve. According to Yaghmai and Maxin,

---

Roy Meador III is at Cumberland College, Williamsburg, Kentucky, and Glenn R. Wittig is on the faculty, School of Library Service, University of Southern Mississippi, Hattiesburg.

> Expert systems are intelligent computer applications that use data, a knowledge base, and a control mechanism to solve problems of sufficient difficulty that significant human expertise is necessary for their solution.[1]

With some exceptions—the word *intelligent* in this context is questionable—this is an acceptable working definition. The *data* include facts about the problem (or domain of problems) to be solved. The *knowledge base* contains clearly defined rules expressed in "if-then" relationships for use in the decision-making process. However, in many systems, heuristics (i.e., intuitive rules of thumb or experience-based guesses) are included. The *control mechanism* coordinates all activities through carefully structured strategies.[2]

These systems require thousands of hours to build, but once completed, they can be updated without having to rewrite the whole program. Because the data, knowledge base, and control mechanism are separate, in contrast to the standard software, they can be altered and updated with a minimum amount of effort. (To aid in this work, there are expert systems available that advise in the construction of expert systems.) The hardest task is that of developing an adequate knowledge base. A majority of the developmental time is spent conferring with human experts about their decision-making processes and developing rules that satisfactorily simulate them.

Suppose, for example, a program were developed to classify two-dimensional shapes into various categories. Such a program would (or should) contain all the known shapes that the experts, who normally perform the classification, would know. In addition the program would have records of all its previous performances of the task and information about its accuracy in performing those tasks. The knowledge base would actually be a working model of the decision-making process in all its intricacies. Included would be the tips or clues the expert would normally use on a subconscious level. The control mechanism would have to be refined to interpret all the rules by running them in their proper sequences. When the specifications of a shape were entered, the program would first search for a perfect match. If one were not found, the program would then search similar shapes or portions of whole shapes, weighing all the possible decisions based on a prioritized list of criteria. The program would check itself against previous "guesses" or, upon the basis of inference formulas, derive probabilities. This could be the first series of steps in a multistep process to present the best classification when no absolute answer could be given. The process could contain hundreds of such routines. Once the "decision" was established, the program could provide the conditions of its "decision making." And if the program were an advanced system, it would allow decisions to be tracked carefully in an online environment where the user could backtrack and alter the course of the process down a different path.

The example above contains the distinctive features of the most advanced systems. Contrary to the optimism exhibited in popular literature, no present system can do all these things. In fact, many believe that the technology is still in a crude state. Boden, for example, described some serious limitations of expert systems:

> They cannot relate their conclusion to the domain in general, or rely on an overview of the problem to assess the relative theoretical reliability of different hypotheses. Nor can they adjust or monitor the structure of their own problem-solving, for they have no high-level representation of it. They are unable, too, to integrate different knowledge domains, and to use concepts and patterns of inference taken from one domain to reason (analogically) in another. Nor can current systems explain their conclusions differently to different users, taking account of the specific user's knowledge. . . . For this reason also, the pattern of interaction between user and system is at present very limited.[3]

Even so, with the present capabilities of expert systems, many practical applications are possible. Since the early 1980s, expert systems have been used outside of research as "consultants" (tools) in hundreds of applications. Research regarding expert systems for library applications, however, has lagged. Machovec, editor of *Information Intelligence, Online Libraries, and Microcomputers*, is optimistic about the number of applications for library automation, but descriptive cataloging is missing from his long list of possible applications.[4] The absence of work in cataloging is glaring, considering that it is the only area with a formalized code. With so many rules in place, it would seem to be relatively easy to develop a knowledge base.[5]

SURVEY OF RESEARCH

Investigations

In 1983 the first publication on the subject of expert systems for descriptive cataloging appeared. Clarke and Cronin devoted a brief section of their article to cataloging and suggested two possible scenarios for research in this field:

- a man-machine interface as a first step towards expert-system cataloging, where the intellectual effort is divided between the intermediary and the support system;
- an expert system with full cataloging capability linked into an electronic publishing system, so that as a text is generated online, it can be passed through an expert-system cataloging process without any intellectual input from the intermediary.[6]

Research in automatic cataloging, which has so far been limited to descriptive cataloging, has followed both of these anticipated routes.

The research carried out by Davies and James in the Exeter Project at the University of Exeter (England) is of the first type.[7] In order to expedite their research because of time and computer memory constraints—to date expert systems have not worked within a domain as large as what the total AACR2 rules would demand—they limited themselves to (1) bibliographic description using templates and (2) two cataloging routines based on AACR2 chapters 21–25 concerning choice of access points and form headings. Their reason for doing the latter was practical. *Using AACR2: A Diagrammatic Approach* by Shaw and others, which represented the rules of these chapters in the form of algorithms, had just been published. A glance at a single page of the book (see figure 1) illustrates both the complexity of the project and the amount of help received from such a representation. To arrive at the point illustrated in the algorithm, a cataloger had to decide that "an entry under TITLE [was] required, [that] a UNIFORM TITLE should be established, and that it [was] a SINGLE TITLE for a work created after 1500."[8]

With such algorithms already available, much of the developmental work on the knowledge base of the expert system was complete. Since their system was not completely automatic and required human input at various stages of the processing, Davies and James designed a menu-driven system, because they believed it to be simple to understand, easy to implement, and possibly faster in application.[9] They chose PROLOG, a high-level programming language, because of its powerful capability for making inferences. After working out a series of technical problems, they came to the following conclusions:

(i) rules for determining access points could be stated as production rules;
(ii) rules governing bibliographic description could be catered for, in many instances at least, through the use of templates;
(iii) local rules regarding the cataloging of certain categories of material could be accommodated;
(iv) menus and screen forms could be designed which would act as a user-friendly interface between a cataloguer and an expert system;
(v) explanation facilities in terms of the system's attempts to decide which rules from AACR2 apply could be easily incorporated.[10]

From the definition of expert systems and the example of how an expert system can operate (both provided above), it is clear that the work of Davies and James in no way approached the full power of an expert system. Although their system had the ability to "leap across" rules of little or no consequence, it lacked flexibility. With the absence of a large knowledge base, it required input throughout the process.

On the other hand, research at OCLC that more fully applies the capability of expert systems is presently following the second route suggested by Clarke and Cronin, i.e., a system with full cataloging capability. Again research is being limited to a particular task, the descriptive cataloging of English language bibliographic items that appear on title pages. To simplify the data and knowledge bases, cataloging will be done on the first level of AACR2. This project, which deals exclusively with automated title-page cataloging, under the management of Stuart Weibel, will investigate the "feasibility of developing an automated cataloging system."[11] The framework of the expert system will be (1) rule

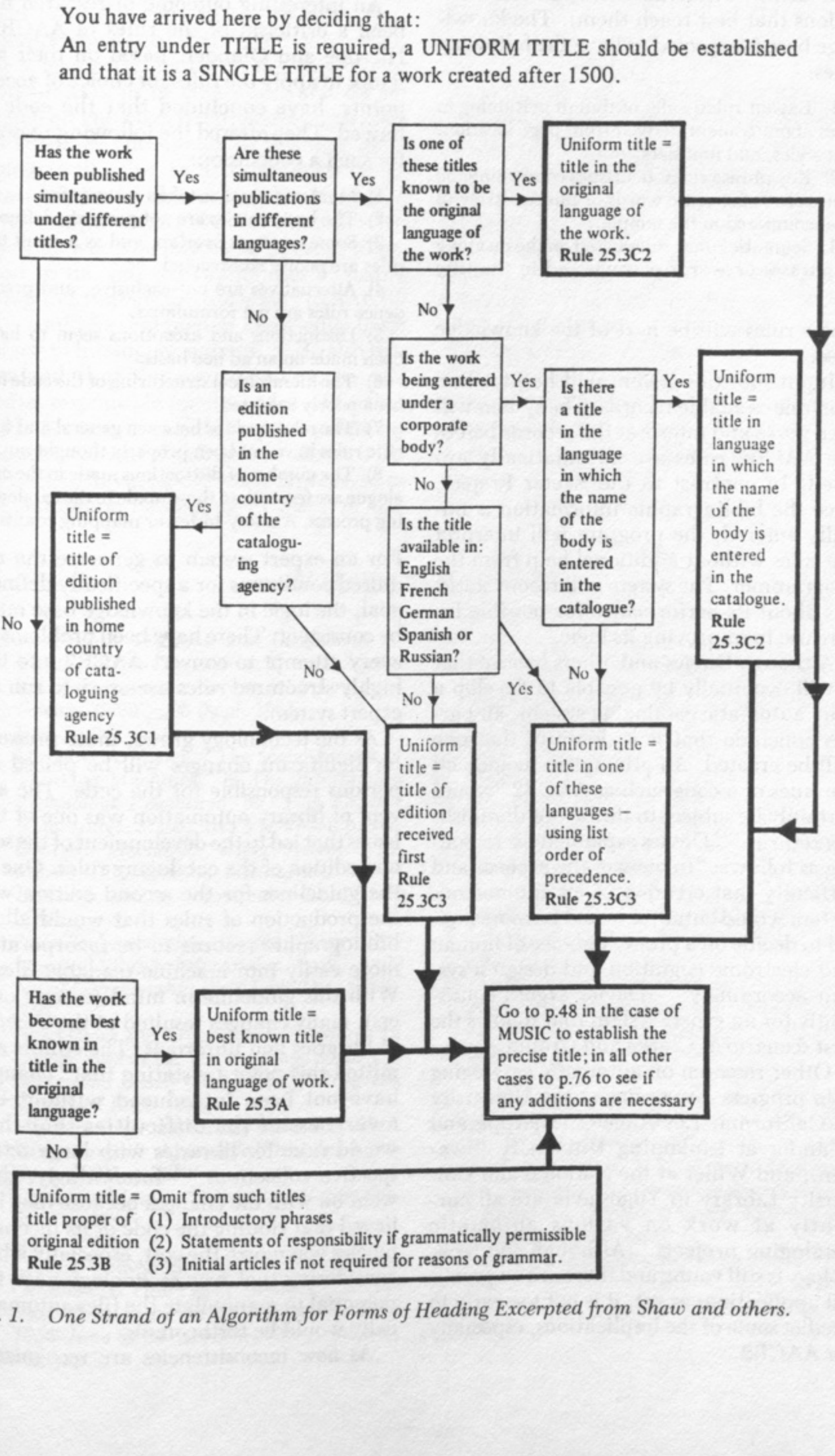

Fig. 1. One Strand of an Algorithm for Forms of Heading Excerpted from Shaw and others.

based and (2) backward chaining (a system that identifies goals and then infers the conditions that best reach them). The knowledge base is expected to have the following rules:

1. Layout rules: rules of thumb pertaining to clues about content derived from page location, font styles, and font sizes.
2. Key phrase rules: heuristics concerning the occurrence of specific words or phrases likely to be encountered in the sample.
3. Semantic rules: rules based on the meaning of phrases or words encountered in the title page.[12]

These rules will be part of the knowledge base.

Input into this system will be based on machine-readable records. The system will then parse and interpret the records before the AACR2 rules are automatically applied. In contrast to the Exeter Project, once the bibliographic information is initially entered, the program will interpret the rules without additional help from the programmer. The system will record statistics about its performance for possible future use in improving its logic.

Although Davies and others believe that it will eventually be possible to develop a fully automatic cataloging system, all parties conclude that it is doubtful that one will be created. An attempt to include all the rules of a code such as AACR2 "would certainly be subject to the law of diminishing returns."[13] Davies explained his reasoning as follows: "In view of the expense and difficulty that creation of an autonomous system would entail, it would be more logical to decide on a proper balance of human and electronic cognition and design a system accordingly."[14] Davies argues consistently for an expert system that follows the first scenario of Clarke and Cronin.

Other research on automatic cataloging is in progress. Svenonius at the University of California, Los Angeles; Hjerppe and Olander at Linkoping University (Sweden); and Willer at the National and University Library in Yugoslavia are all currently at work on various automatic cataloging projects.[15] Although the technology is still young and there are no practical applications as yet, it is not too early to predict some of the implications, especially for AACR2.

Implications

An interesting outcome of research has been a criticism of the rules in AACR2. Hjerppe and Olander, based on their attempt to apply the rules for choice of access points, have concluded that the code is flawed. They offered the following reasons for such a conclusion:

1) Much too much is taken for granted.
2) The basic entries are not properly defined.
3) Some concepts overlap, and as a result the rules are poorly constructed.
4) Alternatives are not exclusive, and precedence rules are not formulated.
5) Distinctions and exceptions seem to have been made on an ad hoc basis.
6) The hierarchical structuring of the code has been poorly executed.
7) The relationships between general and specific rules have not been properly thought out.
8) The number of distinctions made in the catalogue are fewer than those made in the cataloguing process. A many-to-fewer mapping results.[16]

For an expert system to generate the required conditions for a specifically defined goal, the logic in the knowledge base must be consistent. There have been problems in every attempt to convert AACR2 into the highly structured rules necessary to run an expert system.

As the technology grows, more pressure for significant changes will be placed on persons responsible for the code. The advent of library automation was one of the issues that led to the development of the second edition of the cataloging rules. One of the guidelines for the second edition was the production of rules that would allow bibliographic records to be incorporated more easily into machine-readable files.[17] With this guideline in mind (among others), many changes resulted at the expense of libraries and librarians. The editors admitted this point by stating that "changes have not been introduced without the awareness of the difficulties that they would raise for libraries with large retrospective collections."[18] Interestingly, they went on with the changes because they believed that ridding the code of its inconsistencies was worth the cost, especially when considering that new technology with the potential to manipulate the files automatically would be forthcoming.

As new inconsistencies are recognized,

and as the new technology demands flawless logic, there will be less reluctance to hold onto the code in its present form. Hjerppe concluded that "eventually a basis for a new code, based on the capabilities and limitations of computers and on research on knowledge representation should emanate."[19] The impact of this new technology will usher in new problems for librarians, but problems can be minimized with cooperation. Davies agreed with Hjerppe when he said that "in order to exploit to the full the capabilities of future computer systems the successors to AACR2 . . . could be designed as knowledge bases for use in intelligent systems."[20] He also recognized that "encouragement, by the bodies responsible for the codes, of further [AI] research . . . would be prerequisite for the achievement of this goal."[21] There is little hope of having a workable expert system for automatic cataloging in place in this decade. The process is slow, but a system will probably be ready sometime during the 1990s.

CONCLUSION

The literature on expert systems is generally optimistic and sometimes even sensational. Still, many successful, practical applications are currently being used. Unfortunately, there are not many library and information science applications presently available. Current research on automatic cataloging has had mixed results, but researchers continue to suggest further work. The impact of this technology on codes such as AACR2 will be great because of the demand for consistency. When one looks at the previous decisions of those responsible for AACR, it is safe to conclude that more changes in the code will be forthcoming.

REFERENCES AND NOTES

1. N. Shahla Yaghmai and Jacqueline A. Maxin, "Expert Systems: A Tutorial," *Journal of the American Society for Information Science* 35:297–305 (Sept. 1984).
2. Ibid., p.297–98.
3. Margaret A. Boden, "The Social Impact of Thinking Machines," in *The Information Technology Revolution*, ed. Tom Forester (Cambridge, Mass.: MIT Pr., 1985), p.101.
4. George S. Machovec, "Expert Systems in Libraries," *Information Intelligence, Online Libraries, and Microcomputers* 2:1–4 (Dec. 1984).
5. Ann Clarke and Blaise Cronin, "Expert Systems and Library/Information Work," *Journal of Librarianship* 15:277–92 (Oct. 1983).
6. Ibid.
7. Roy Davies and Brian James, "Towards an Expert System for Cataloguing: Some Experiments based on AACR2," *Program* 18:283–97 (Oct. 1984).
8. Malcolm Shaw and others, *Using AACR2: A Diagrammatic Approach* (Phoenix, Ariz.: Oryx, 1981), p.75.
9. Davies and James, "Toward an Expert System," p.288.
10. Ibid., p.295.
11. Online Computer Library Center, Inc., "Automated Title Page Cataloging," in *Annual Review of OCLC Research: July 1985–June 1986* (Dublin, Ohio: OCLC, 1986), p.9.
12. Ibid.
13. Roy Davies, "Cataloguing as a Domain for an Expert System," in *Intelligent Information Systems: Progress & Prospects*, Artificial Intelligence Series, ed. Roy Davies, (New York: Wiley, 1986), p.69.
14. Ibid.
15. Id., personal letter, November 3, 1986.
16. Roland Hjerppe and Birgitta Olander, *Artificial Intelligence and Cataloging: Building Expert Systems for Simple Choice of Access Points: Results and Revelations* (Linkoping, Sweden: LIBLAB Dept. of Computer and Information Science, Linkoping Univ., June 1985), p.17–18.
17. American Library Association and others, *Anglo-American Cataloguing Rules*, 2d ed., ed. Michael Gorman and Paul W. Winkler (Chicago: ALA, 1978), p.vii.
18. Ibid., p.x.
19. Roland Hjerppe, "What Artificial Intelligence Can, Could, and Can't Do for Libraries and Information Services," in International Online Information Meeting, *Proceedings of the 7th International Online Information Meeting* (Oxford: Learned Information, 1984), p.16.
20. Davies, "Cataloguing as a Domain," p.75.
21. Ibid. ■■

Special Section: Measuring System Performance

Editor's Note: At the 1987 ALA Annual Conference, the Technical Standards for Library Automation Committee (TESLA), a committee of LITA, presented a program entitled "Measuring System Performance: A Challenge for Libraries and Systems Providers." This program addressed the possible role of standards in measuring the performance of automated library systems and in planning the needed capacity of a system. The following articles are based upon presentations given at that program. Special thanks are due to Dorothy McPherson, past chair of TESLA, and Marilyn Nasatir, the current chair, for organizing this program and for having the papers submitted to ITAL.

Capacity Modeling at RLG

James G. Schmidt and Michael Pobuda

The model developed by the Research Libraries Group (RLG) to project resource utilization is presented here because it has worked for us; it is not our intention that you should hunt for its virtue or, alternatively, be stunned by its imperfection. It seems to do for us what we need a model of this kind to do, and we are pleased to share it. Before describing the model, we would like to make three points about the context in which it was created.

The first point is that as a practical matter we, like those of you who are installing local systems, had a need to develop something that would allow us to plan and project the CPU requirements for a given period of time to support a given level of activity.

Second, we sought something that would generate the evidence needed for our board's authorization of the sizable capital expenses associated with upgrading CPU capacity. Happily, that threshold of capital expense is buying us steady improvements in price performance.

The third point concerns the elements or concepts for the model we use. Those that are essential to a capacity model seem to us to be as follows: (1) the kinds of transactions to be processed, (2) the number of transactions to be processed, (3) the amount of resource required by type of transaction, and (4) the varieties in the mix of transactions over time.

TYPES OF ACTIVITY

The first thing we did at RLG was to figure out what types of activities were taking place on the system and then to count them. In the technical processing system, cataloging, acquisitions, retrospective conversion, searching, and interlibrary loan were included. How many records were being entered? How many items were being cataloged? How many searches were being made? Then the question became, Can it be counted? Can you count how many searches are coming through? Does the system provide that sort of statistic? Once we had done that, we began to look at the different methods of accounting that the system provided for us and to develop some of our own.

TYPES OF RESOURCES

We then examined the specific resources being used by each transaction. How much CPU or IO is required for a cataloging transaction? How much CPU or IO is needed for a typical search? Development

---

James G. Schmidt is Vice-President and Director of RLIN, and Michael Pobuda is a Systems Programmer at the Research Libraries Group in Stanford, California.

work is not countable in the same way that user transactions are but is countable in the sense that staff tend to use a certain amount of CPU and a certain number of IOs per day. On the other hand, cataloging work is countable: we count the number of records entered into the system, the number of records entered by using the acquisitions system, or the number of ILL transactions.

MEASURING RESOURCE UTILIZATION

It is important to measure the current level of use and from that, to project. If you get into a crisis mode where response time is becoming slower and slower, you may not be able to understand the problem unless you are already tracking CPU utilization, number of IOs, etc.

At RLIN, we have a set of measurement tools. First of all, RLIN runs on a MVS system, which provides a facility called Resource Measurement Facility (RMF). It writes records, detailing the amount of CPU and IO that major subsystems use. For us, the major subsystem is the Stanford Time Sharing System (STS), which provides all of our time-sharing facilities and also writes its own accounting records, detailing the resource consumed in each user session. Finally, the technical processing system itself logs a certain percentage of sessions; 5 percent of all sessions are logged on a per-command basis so that we can statistically detail the CPU and the average IO rate for a particular command. For each session and for each major subsystem, then, we know how much CPU resource is being used.

PROJECTING FUTURE USE

We begin each year by developing use projections for the following year. These are based on a survey of our members, who tell us how many cataloging transactions they expect to do, how many items they expect to push through the acquisitions system, etc., so that we can create a business plan for the next year. Then, looking at the detailed statistics that we have gathered about CPU and IO usage, etc., we can project our CPU requirements for the next year on a per-transaction basis. Figure 1 shows a fairly detailed printout of the FY87 CPU model that was prepared with a PC tool called Javelin. It is somewhat similar to the 1-2-3 spreadsheet but is more useful for analyzing time-series data. We take the full-year projections gathered by our users and spread them over the entire year by looking at the past three years. The number of transactions in any given month tends to vary, depending on the period of the academic year, and there is a definite pattern. After spreading the use over the year, we then develop projections based upon budgeted and maximum activity.

The budgeted activity represents what is expected, and the maximum is a projection of the worst case. We combine those activity numbers with the resource numbers on a per-transaction basis to arrive at a projection of utilization (see figure 1). In addition to projections based upon expected and maximum numbers, we also make them for the peak periods. For RLIN, there is a definite peak period, usually between 8 a.m. and 2 p.m. (Pacific time), when libraries in all time zones are using the system.

The projections for resource utilization are broken down into several components, as shown in figure 1. Orvyl, the major component of STS, consumes the largest proportion of our resources and is where all the online work is done in the technical processing system. HiPriority includes tasks that are essential for running the network. Batch includes low priority tasks associated with development of new systems. MVS shows the overhead required for the MVS Operating System. Aspen and MDF reflect development time spent for our new operating system environment.

PROJECTING FUTURE NEEDS

We add these numbers up and ask, Will it all fit on the existing machine? As figure 1 shows, for budget utilization, it does fit. For peak utilization, that is to say, the busiest times of the day, it will fit most of the time, but notice that in February we will exceed our existing CPU by 4 percent. Several of the following months will also require more CPU. Looking at the projections for the maximum or worst case, you will see that we start to exceed the capacity of the machine by a substantial amount. It was this model that caused us to upgrade

13 Aug 1986 FY87 CPU Model - Constants

	Cat	CatMai	Recon	Acq	AcqMai	ILL	Print
CPU per BU	0.583	0.298	0.492	0.592	0.373	1.643	0.074

	Wylbur	Milten	Jes2
Day Utilization	4.1%	2.4%	0.6%
Peak Utilizatio	4.7%	2.9%	0.5%

	Orvyl	MVS
Day Overhead	76.9%	21.1%
Peak Overhead		19.6%

Peak to Average Ratio
127%

13 Aug 1986 FY87 CPU Model - Activity

	Sep 1986	Oct 1986	Nov 1986	Dec 1986	Jan 1987	Feb 1987	Mar 1987	Apr 1987	May 1987	Jun 1987	Jul 1987	Aug 1987	1987
fy87 budget activity													
catalog	124,929	146,056	125,985	112,511	140,825	143,306	148,039	161,614	156,849	154,848	158,273	149,707	1,722,942
catmai	124,901	144,269	124,414	106,835	134,018	138,305	145,046	158,725	149,395	148,608	161,908	164,459	1,700,883
recon	49,654	64,610	59,673	47,087	59,344	68,675	69,437	85,933	82,538	78,848	84,496	78,939	829,235
acquisitions	46,909	57,284	48,512	40,316	49,485	54,353	49,148	54,027	55,972	52,084	54,230	46,838	609,159
acqmai	55,770	64,684	55,665	47,077	61,542	64,467	64,254	71,125	69,180	67,100	67,634	57,329	745,829
tp	402,163	476,904	414,250	353,827	445,214	469,106	475,925	531,425	513,935	501,488	526,541	497,272	5,608,048
avg defq	19,151	20,735	23,014	16,849	21,201	24,690	21,633	24,156	25,697	22,795	23,934	23,680	22,294
ill	14,337	19,828	18,201	11,545	14,191	20,442	19,502	19,856	16,030	16,773	19,253	16,825	206,783
fy87 max activity													
catalog	132,815	155,276	133,938	119,614	149,715	152,352	157,384	171,816	166,750	164,623	168,264	159,158	1,831,705
catmai	134,936	156,055	134,689	115,488	144,883	149,748	157,005	172,080	161,944	161,024	175,540	178,278	1,841,670
recon	55,794	72,600	67,052	52,910	66,683	77,167	78,024	96,560	92,745	88,598	94,945	88,701	931,780
acq	51,319	62,670	53,073	44,106	54,137	59,463	53,769	59,107	61,234	56,981	59,328	51,241	666,428
acqmai	61,014	70,765	60,899	51,503	67,328	70,527	70,295	77,812	75,684	73,409	73,992	62,718	815,946
tp	435,877	517,366	449,650	383,621	482,746	509,258	516,478	577,374	558,358	544,635	572,070	540,096	6,087,529
avg defq	20,756	22,494	24,981	18,268	22,988	26,803	23,476	26,244	27,918	24,756	26,003	25,719	24,201
ill	15,464	21,386	19,631	12,452	15,306	22,049	21,035	21,416	17,290	18,091	20,766	18,148	223,034

1

13 Aug 1986 FY87 CPU Model - Utilization

	Sep 1986	Oct 1986	Nov 1986	Dec 1986	Jan 1987	Feb 1987	Mar 1987	Apr 1987	May 1987	Jun 1987	Jul 1987	Aug 1987	1987
budget utilization													
Orvyl	42.2%	45.4%	49.3%	38.5%	45.3%	51.7%	46.4%	50.1%	52.0%	47.5%	49.8%	49.0%	47.3%
HiPriority	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%
Batch	7.5%	7.5%	7.5%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.6%
MVS	12.0%	12.6%	13.4%	10.7%	12.1%	13.4%	12.3%	13.1%	13.5%	12.6%	13.0%	12.9%	12.6%
Aspen & MDF	5.0%	5.0%	5.0%	5.0%	10.0%	10.0%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	9.6%
Total	73.7%	77.6%	82.2%	66.2%	79.4%	87.2%	83.2%	87.7%	90.0%	84.6%	87.4%	86.4%	82.1%
budget peak utilization													
Orvyl	53.6%	57.7%	62.6%	48.9%	57.6%	65.7%	58.9%	63.6%	66.1%	60.4%	63.3%	62.2%	60.1%
HiPriority	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%
Batch	7.5%	7.5%	7.5%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.6%
MVS	13.6%	14.4%	15.4%	12.2%	13.9%	15.5%	14.1%	15.1%	15.5%	14.4%	15.0%	14.8%	14.5%
Aspen & MDF	5.0%	5.0%	5.0%	5.0%	10.0%	10.0%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	9.6%
Total	87.9%	92.8%	98.6%	79.2%	94.6%	104.3%	98.7%	104.3%	107.3%	100.4%	104.0%	102.7%	97.9%
max utilization													
Orvyl	49.1%	52.6%	56.8%	45.1%	52.4%	59.4%	53.6%	57.7%	59.8%	54.9%	57.4%	56.5%	54.6%
HiPriority	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%
Batch	10.0%	10.0%	10.0%	5.0%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.9%
MVS	13.9%	14.7%	15.6%	12.0%	14.1%	15.6%	14.4%	15.2%	15.7%	14.6%	15.2%	15.0%	14.7%
Aspen & MDF	5.0%	7.5%	10.0%	10.0%	12.5%	12.5%	15.0%	15.0%	20.0%	20.0%	20.0%	20.0%	14.0%
Total	85.0%	91.8%	99.4%	79.1%	93.6%	102.0%	97.5%	102.4%	110.0%	104.0%	107.1%	106.0%	98.2%
max peak utilization													
Orvyl	62.4%	66.8%	72.1%	57.3%	66.6%	75.5%	68.2%	73.3%	76.0%	69.7%	73.0%	71.8%	69.4%
HiPriority	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%	8.2%
Batch	10.0%	10.0%	10.0%	5.0%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.9%
MVS	15.3%	16.2%	17.2%	13.8%	15.7%	17.4%	16.0%	17.0%	17.5%	16.3%	16.9%	16.7%	16.3%
Aspen & MDF	5.0%	7.5%	10.0%	10.0%	12.5%	12.5%	15.0%	15.0%	20.0%	20.0%	20.0%	20.0%	14.0%
Total	100.8%	108.6%	117.5%	94.2%	110.4%	121.1%	114.8%	120.9%	129.1%	121.7%	125.5%	124.1%	115.7%

Fig. 1. FY87 CPU Model.

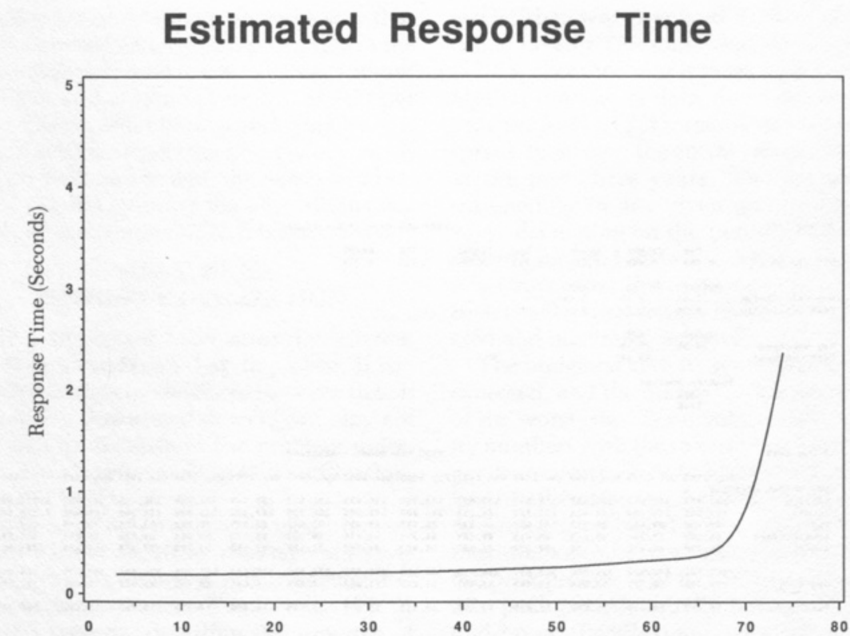

Fig. 2. Projected Orvyl Utilization.

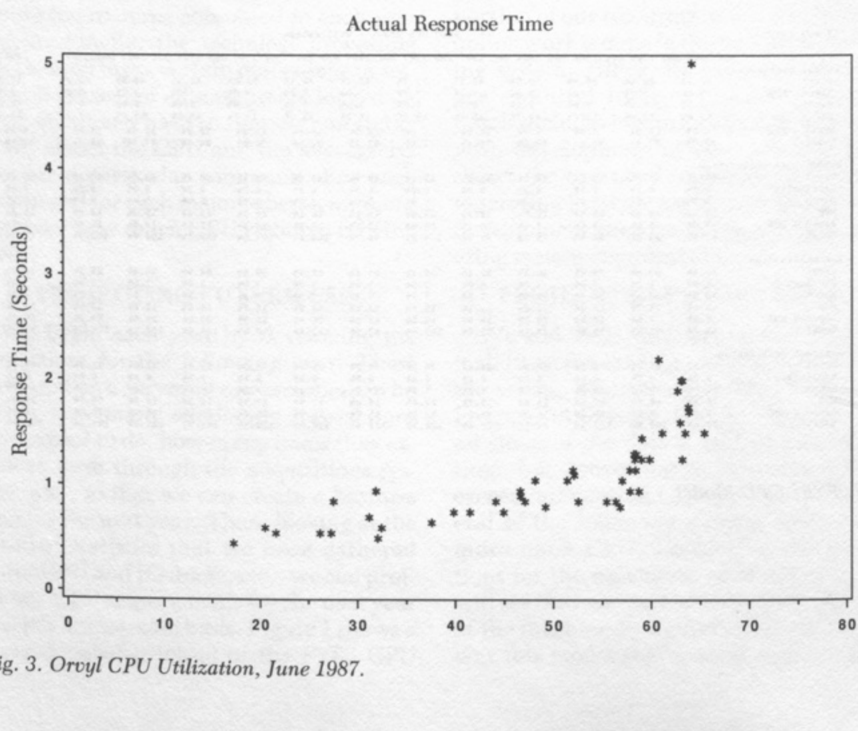

Fig. 3. Orvyl CPU Utilization, June 1987.

our machine earlier this year. It allowed us to predict a need for "more machine"; we began planning, asking vendors for estimates, and scheduling delivery.

The value of a capacity plan is that you can anticipate instead of letting it happen to you. Without such a plan, you may be scurrying around at the last minute to find the money and other resources for upgrading the machine. You may not have budgeted for a new machine, or you may not have space for one. It is essential to know in advance what you will need to meet demand.

VALIDATION

The last major part of our resource utilization model is validation. Was the model correct? Do we see resource utilization as predicted? Figure 2 is an estimated response-time graph, which projects utilization on the machine. It comes from a simple queueing model and predicts service time—that is to say, response time—of the CPU as Orvyl utilization grows. Figure 3 shows some actual data from June 1987. The graphs in figures 2 and 3 match up fairly closely, which validates our model.

It is essential to return to your projections, compare them with actual data, and ask, Did we realize the response time predicted? Is the CPU pegged as we predicted, or is something missing? If your projection does not match the actual utilization, you must go back and identify the problem—perhaps some small amount of system overhead, perhaps some group of transactions that was ignored earlier. It will be necessary to feed this additional information into the model, run it again, get a new projection, and check that against reality. ■■

Response Time Measurement and Performance Analysis in Public Access Information Retrieval Systems

Clifford A. Lynch

This paper discusses both the measurement of response time and related performance data in large, interactive, public access information retrieval systems such as online catalogs and the uses such data may have in managing and tuning this type of retrieval system. Although actual experience with the University of California's (UC) MELVYL[1] online union catalog provides the majority of the examples in the paper, most of the points made are equally applicable to other locally developed or vendor-supplied systems. However, many vendor-supplied systems provide only minimal performance and usage-monitoring data collection facilities, making the information discussed here unavailable from turnkey systems. Similarly, an installation's options for tuning a turnkey system may be quite limited. This lessens, but does not eliminate, the value of recording and analyzing detailed performance data if this data is actually available for collection in a turnkey environment.

It is important to maintain a perspective on performance measurement that focuses on the goals of the effort. Richard Hamming states this eloquently: "The purpose of computing is insight, not numbers."[2] System performance is measured to understand how well the system is working and, ideally, how to improve its performance. This purpose is easily obscured in a flood of detailed performance data or in weekly charts of a few mysteriously derived values that allegedly characterize system performance. High-level management, pressured to improve the performance of a troubled system, may become obsessed with using a few arbitrarily selected numbers to demonstrate that this week's performance is *better* than last week's or that system response time is *satisfactory*. This is a particular danger when the managers in question have a

Clifford A. Lynch is Director, Division of Library Automation, University of California System, Berkeley.

background in quantitative analysis and are searching for simple "scientific" answers to complicated, poorly posed, and somewhat subjective problems.[3] System performance is not characterized honestly by a few numerical values. One of the greatest difficulties challenging system managers of large, complex public access information retrieval systems, where the system services a wildly varying set of searches from one day to the next, is the identification of meaningful measures of system behavior.

In the analysis of performance data, it is essential to look at a wide range of measurements and not just one or two selected numbers. For example, a system can run far below capacity, yet still provide very poor response time. If most of the elapsed time in the processing of a user search is spent waiting for input/output operations of some kind, or for control over some serialized resource, upgrading to a faster machine may fail to improve response time significantly because all systems wait at the same speed. (A faster process, however, may allow the system to support a larger number of concurrent users before response time gets any worse.)

Examination of the low CPU utilization might lead an analyst to conclude that the system is not only running efficiently but that it can be expanded. Obviously, this conclusion would be wrong under the conditions just described.

The analyst must constantly fight myopia and preconceptions. In most real systems, performance is limited by a small number of bottlenecks at any given time; however, when one is eliminated, a new one will limit system performance. Identifying a key bottleneck may require protracted and persistent detective work. Once found and eliminated, the analyst may continue to monitor the eliminated bottleneck, believing that because it has been successfully eliminated the entire performance problem is solved. In practice, elimination of a single bottleneck (except in the early stages of a system's life cycle, when it is still poorly tuned) will almost always have less overall effect than the person eliminating it would like to believe.

GOALS OF PERFORMANCE MEASUREMENT

If the primary purpose of performance measurement is to gain insight, the starting point for a discussion of performance measurement should be a review of the types of insights one hopes to obtain. There are four major (and interrelated) areas in which performance data are useful:

• Capacity planning. In this application, the amount of existing resources consumed to support use at a given level (for example, a certain number of concurrently active terminals) is calculated, and extrapolation is then used to project when the system capacity will need to be enlarged. Capacity planning is an essential part of the budgeting processes.

Capacity planning usually involves a good deal of inspired guesswork. The problem is to extrapolate a function based on a limited number of known data points. This resource utilization function is usually complex and contains nonlinear factors. Sometimes such elaborate validation techniques as system modeling and simulation are used to attempt to support resource utilization extrapolations.

• Acceptance testing. A new or upgraded system is often installed as part of a contract that specifies certain performance requirements. Clearly, one wants to verify that the system does meet the performance requirements specified in the contract before accepting it and agreeing to pay for it.

Acceptance testing often has legal ramifications. The contract is usually not precise enough to translate directly into implementable measurement specifications, thus introducing the need for interpretation that may be contested by the vendor. Acceptance testing may also take place in an atmosphere of intense administrative pressures urging the analysts to prove that the system meets specifications (or, occasionally, that it fails to meet specifications). Insight into system performance is not always on the agenda during a system acceptance period.

• Quality assurance. During a system's lifetime, its performance may degrade for any number of reasons, ranging from database growth to shifts in the way the user community uses the system's facilities. Ongoing performance measurement ensures that these problems become visible to the system management, hopefully at an early stage, so that they can be corrected before a major response-time crisis arises and the user community begins to lose confidence in the system.

Unfortunately, as with acceptance testing, quality assurance may take on strong political overtones. There may be a mandate to the analysts to prove that the system is working well (or performing poorly) so that the analyst's conclusions can support management's long-term system expansion or replacement plans.

• System tuning and refinement. The amount of useful work that a site can perform in this area will vary, depending on the particular system.

At one extreme, the library with a simple turnkey online catalog may be able to identify areas of the system that need improvement but may have no tuning facilities and have to rely on the vendor for any kind of tuning or incremental design improvements. Even in this environment, however, knowledge of system weaknesses is valuable to the library in designing programs to train patrons to use the system efficiently.

At the other extreme, a large, locally developed retrieval system running on a large mainframe offers a great variety of tuning opportunities. One can tune parameters and configuration options in the operating system, in the telecommunications and networking software, or in the database management system (if a DBMS is used). For example, an entire shelf of manuals is needed simply to document tuning options available in IBM's top-of-the-line MVS operating system. As well as tuning the environmental software, the application can be tuned to work more efficiently in its operating environment by adjusting factors ranging from the placement of data on disk to the invocation of system supervisor services in the applications code. Finally, detailed performance data can be a major resource in guiding the redesign and evolution of the user interface in an information retrieval system.[4]

Quality assurance can be viewed as an ongoing acceptance testing program and requires the same sorts of information to be gathered. System tuning and capacity planning, however, generate rather unique needs for data. There are three basic types of data that can be collected:

• Statistical data. This primarily involves counting and averaging events. Typical statistical measures include the number of searches run per month, the number of records displayed, the number of connect hours of service provided, and the average response time for certain types of commands. Statistical data are useful in all types of performance measurement and analysis, but although it is inexpensive to collect, it is surprisingly hard to validate and present intelligibly.

• Benchmark data. Here one develops a script of system requests, runs it, and measures the time required to execute the script. Sometimes resources consumed in executing the script are also measured. This method provides a way of comparing system performance at one point in time with performance at another point in time. It is quite difficult, however, to develop scripts that accurately reflect *typical* user work load; consequently, benchmarking is of limited use in ongoing quality assurance, for example, unless the validity of the scripts used is continually reassessed against actual searches entered by users.

• Detailed logging data. One logs detailed information about every command performed by the system or every activity of some type (such as each module call or I/O operation). Most large, locally developed systems generate this sort of data, at least optionally, to help with system tuning and development. Current commercial operating systems offer this level of detail about the low-level hardware and software systems as an option that can be enabled and disabled. There is a wealth of useful information and potential insight in detailed logging data; however, due to the huge volume of data generated at this level of detail, it can be tremendously difficult to use. Typically, the site must build or purchase a variety of analysis tools to extract the insights hidden in this mass of data.

All three types of performance data can be measured either from an applications or a computer systems perspective. Considering applications, response time (discussed below) is the key measurement. For the computer systems perspective, resource consumption (I/O operations, CPU utilization, and memory utilization) is the most commonly measured area.

COMPUTER-SYSTEM-ORIENTED MEASURES

Keeping track of resource consumption is important for system tuning. Applications functions that consume substantial resources can be identified and modified to reduce their resource consumption.

Resource consumption measurement is also the key to capacity planning in a traditional computing installation. However, conclusions drawn from resource consumption data in a sophisticated information retrieval system routine may be inaccurate. Consider the case of a stand-alone computer that supports a large online catalog. The machine is a fixed resource. User load on the system fluctuates during the day. A primitive retrieval system executes user queries without regard to system loading, and, in this case, resource utilization is correlated to system load. In comparison, a very sophisticated system may execute queries using different strategies based on the overall level of system load. When the system is heavily loaded, the system seeks to minimize the resource consumption of individual queries, even if this leads to longer execution time for each query. In light load situations, the system may employ resources that are otherwise idle to minimize elapsed execution time for queries and, thus, the response time perceived by the user. Techniques such as precomputation, parallel computation, prefetching of data, and extensive caching of data trade computing resources for better response time.[5] As retrieval systems become more sophisticated, use of load-sensitive, multi-objective optimization techniques will increase.

In a system that employs such advanced optimization methods, computer-system-level resource utilization information may contribute little to capacity planning. Even under fairly light loads, the system should display a high degree of computing resource utilization. Capacity planning in this environment requires application-oriented data, such as average response time under peak loading.

Other extraneous factors can also confuse the interpretation of system-oriented performance data, even if the retrieval system employs simplistic query optimization methods. Suppose the retrieval system runs on a computer that functions as a service bureau and recharges for running the application (for example, an online catalog that runs at a commercial service bureau or a campus computer center that recharges the library in an academic setting). In this environment, the simplistic approach to query optimization is beneficial because the computing resources are not a fixed cost commodity. In addition, the application may be designed to minimize the recharge from the service bureau. Service bureau recharge algorithms vary greatly from one computer center to another: some charge heavily for I/O and very little for CPU cycles; others penalize heavy memory use and CPU consumption while ignoring I/O charges. Minimizing recharge cost may have little to do with system performance optimization and may actually distort capacity planning efforts; yet it may become the primary management objective in such a recharge environment.

APPLICATION-ORIENTED MEASUREMENT DATA

The simplest type of application-oriented data is statistical: one can count the number of events such as searches run, sessions conducted, records displayed, and help screens viewed. A small number of such counters can provide a rough sense of load, which is useful in capacity planning. As this data is applied to system refinement, however, the number of events that must be counted multiplies rapidly, and each event becomes precisely defined (for example, searches that are not rejected due to a syntax error, searches that use a single index, and searches that fail to identify any records in the database).

The other key type of application-oriented performance data is response time, and it really is the most important single measure of system performance in a public access system. Such systems live or die on response time and reliability.

Response time is a difficult measure to define (even ignoring the pragmatic problems of collecting the measurements). Imagine a person equipped with a stopwatch sitting at a system terminal. A command is entered, and the stopwatch starts ticking. When should the stopwatch be halted? The general answer is "when the system responds." But turning this into a specific definition creates many problems. Consider, for example, a system supporting a search command that normally reports the number of records matching the search criteria and a display command that lists records in the search result. Consider these system details:

• Many online catalogs automatically display the search result when the search identifies a single record or a small number of records. Some searches measure search response, and other searches measure the time for a combined search and display operation.

• The system may follow generally accepted user interface design practices by not letting the user wait too long for a response; while processing a search, it may periodically issue such messages as "The system is processing your request." Is this message a search response?

• The system may respond to a search with a display such as "Your search will retrieve approximately 22,000 records" and then offer the user the following menu of choices:

1) Enter "continue" to continue the search.

2) Enter a new search.

3) Enter "sample" to see 100 sample result records.

4) Enter "analyze" to have the system suggest ways to reduce the size of the expected search result.

Is this a search response? What happens if the user enters "continue"? How should one account for the possibility that the user will sit thinking for a few minutes before making a choice? As the user interface becomes more sophisticated and guides the user more—offering choices and trying to presume the user's needs—defining both a transaction and a useful measure of response time becomes more difficult.

• Measuring display response time might seem simpler: halt the stopwatch when the record is completely displayed. Display should be a quick operation and should not require any user interaction or interim system messages. However, if the record to be displayed is 800 characters and if the user's terminal runs at 300 baud (for example, a low-end microcomputer dialing up over a phone line), it will take about 30 seconds to display the records. If the terminal is running at 9600 baud (for example, a hard-wired terminal in a library), then it will take less than a second to physically write the record on the terminal screen. The actual measured display response time may be almost entirely determined by the characteristics of the terminal used to access the system.

In practice, these problems are usually solved by rather arbitrary decisions. Assume now that response time has been defined somehow and data on response time can be collected. Interpretation of this data is also treacherous. In a transaction-oriented application such as circulation, it is reasonable to assume that if the system is working properly, each operation—such as checking in a book—should take about the same amount of time and consume about the same amount of resources. In a public access searching system such as an online catalog, this assumption is no longer valid. The variation in the amount of computational work performed from search to search is enormous. In the MELVYL system an average search result is about 100 books. However, about a quarter of the searches find no books at all, and the standard deviation of the search result size is over one thousand! This means that average response time is a rather meaningless figure. The measurement that should really be used is some ratio of the amount of time that the command takes running standalone on the system to the amount of time the command actually takes in practice. This figure gives some indication of response time degradation due to load. If this ratio remains within acceptable range, and if performance for user queries is satisfactory when the system is not under load, then the system is performing satisfactorily. However, to compute the ratio, the search must be run on an idle system to establish the baseline execution time. This may be unfeasible, and at best it precludes any attempt to measure system performance in real time.

It is possible to use the ratio method in benchmarking, if one can devise a script that accurately reflects the full range of *typical* system behavior. One runs the benchmark stand-alone and measures response time, which sets up a baseline. Later, the script is rerun while the system is under load, and the degradation ratio is computed.

Response time also suffers from nonrepeatability. Since response time is closely related to database size, as the database grows, commands may take longer because they retrieve more material. This can even be a problem when a benchmark script is run periodically. As the database grows, the elapsed time for various commands may scale up, probably in a nonuniform and nonlinear fashion.

Response time data is as treacherous to collect as it is to define and analyze. The simple approach is to put code into the online catalog application that time-stamps a command on entry to the system and then computes the difference between current and entry times to determine response time for the command as it writes out a response to the user's terminal. Unfortunately, this may produce completely fallacious response time data. Consider the process of writing a command response line to a user in the MELVYL catalog: the application program passes the response to a terminal handling program; the terminal handler passes the response to the networking software; the networking software passes the data to the operating system for transfer from the CPU to the local network interface. All of these activities add delay to the delivery of the response, yet all of them are invisible when response time is measured by the applications program. In theory, one wants to measure entry and exit times for a command as close to the "outside" of the CPU as possible, most likely in the networking software. However, the networking software is a general-purpose package; it has no knowledge of what constitutes an applications request and what constitutes a response from the application.

Delay generated by multiple software components within the CPU is not the only factor in response time. Once the response actually leaves the CPU, it passes through a telecommunications network to reach the user's terminal. At UC, access to the MELVYL system can be made from anywhere in the Internet, whether from a workstation attached to a local area network at a nearby campus linked directly by a land line or a terminal in Europe that must be reached by passing across five networks and over a transmission path through a geosynchronous satellite. Even worse, the Internet performs dynamic adaptive routing, so network latency may fluctuate wildly from minute to minute. During periods of Internet instability (which can be caused by many events, including the process of reconfiguring routing and connectivity after a major trunk or getaway failure or the onset of transient network congestion), network transmission times can leap from a second or two to half a minute.

Having enumerated the problems of measuring response time, the goals must be reconsidered. If the objective is to perform quality control, the simplest solution is to place external computers (typically microcomputers) emulating actual users at user sites to gather response time data. These external computers will see exactly what a real user sees, including all the aggregate delays caused by the layers of software in the host and the network. However, if the goal is a detailed understanding of the cause of unacceptable response time, a much more complex monitoring system is necessary. The first step is to determine the contributions each component makes to response time. For the network, this requires a clock in each monitoring microcomputer that is closely synchronized with the mainframe.[6] Then exit time at the mainframe can be compared with arrival time at the microcomputer to calculate delay caused by the telecommunications network. Contributions to delay from the various software components in the mainframe can be determined by detailed analysis of application- and system-level event-by-event trace data.

CONCLUSIONS

The key measurement is response time as perceived by the user. This is best measured by an external system that actually emulates a human being at a terminal. Such data is essential for acceptance testing quality control and provides an essential starting point for all further analysis of performance data.

While such a monitor supplies the raw numbers, it does not automatically provide insight about acceptable response time from the user's perspective. This is a highly subjective matter, based as much on the user's beliefs about the complexity of commands (in other words, the user's expectation of how long the command should take) as on actual measured response time.

One of the most important activities in validating a new system design is establishing standards for satisfactory response time that agree with the user's subjective expectations. Good interface design should minimize the cases in which system behavior surprises the user. A badly designed interface will almost guarantee ongoing complaints about response time problems from users. A well-designed interface allows the definition of achievable response time targets that can then be used in an ongoing program of quality control through performance monitoring.

Other measurements, while they may be of interest to the system designer, refiner, or capacity planner, are really secondary. They can be used to solve problems that appear in the measurement of overall response time. While these other measurements may be used to *prove* all kinds of conclusions about the behavior of the system, the conclusions may be illusory if not reflected in actual response time as perceived by the end user.

ACKNOWLEDGEMENTS

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

REFERENCES AND NOTES

1. The registered trademark of The Regents of the University of California.
2. Richard Hamming, *Numerical Methods for Scientists and Engineers* (New York: McGraw-Hill, 1962).
3. Recent history provides many examples of the misuse of quantitative methods in both business and government. Two books by David Halberstam, *The Reckoning* (New York: Bradford Morrow, 1986) and *The Best and the Brightest* (New York: Random, 1972), provide illuminating case studies of this process.
4. Incremental system redesign based on information obtained from system monitoring is discussed in Clifford A. Lynch and Robert K. Brandriff, "The Evolution of the User Interface in the MELVYL Online Catalog, 1980–1985," *Proceedings, 48th ASIS Annual Meeting, Las Vegas, Nevada, October 20–24, 1985* (White Plains, N.Y.: Knowledge Industry, 1985), p.102–5 and particularly Clifford A. Lynch, "Applications of Performance and Usage Data for Online Catalogs," in *Online Library Catalogs: Essential Questions and Future Prospects*, ed. Charles Hildreth, forthcoming.
5. See Clifford A. Lynch, "Extending Relational Database Management Systems for Information Retrieval Applications" (Ph.D. diss., Univ. of California-Berkeley, 1987) and Clifford A. Lynch, "Developments in Database Management System Technology and Their Impact on Information Retrieval," submitted to 51st ASIS Annual Meeting, 1988.
6. The maintenance of a global clock (i.e., a set of synchronized clocks at each node) in a distributed system is a crucial problem. This problem can be solved either by purchasing extremely expensive clocks that will remain synchronized for long periods of time to a common time standard (for example, cesium atomic clocks); by synchronizing all clocks to a common external reference such as the National Bureau of Standards WWV shortwave clock signal; or by doing adaptive clock synchronization using algorithms that exchange data over the network. See David Mills, "Experiments in Network Clock Synchronization," ARPANET RFC 957, and "Algorithms for Synchronizing Network Clocks," ARPANET RFC 956, available from DDN Network Information Center, 333 Ravenswood Center, SRI International, Menlo Park, CA 94025. ■■

Measuring System Performance: The Library's Perspective

Julie Brown

Stephen R. Salmon, Carlyle Systems; Susan Baerg Epstein, consultant; and staff members from Boston University were all part of a stress/capacity test conducted at Carlyle headquarters in both March and April 1986. Although perspectives will differ, our experiences can prove useful to others contemplating such a test.

Stress tests are aptly named: they stress the vendor, the customer, and the system and should not be undertaken lightly. If the vendor has an installed site with a working system of the size and complexity needed, that is the best way of judging the suitability of the system. When that is not the case, a stress test is the only way to find out if a system meets the customer's requirements.

WHY STRESS TESTS?

Boston University decided, for several reasons, to add to its contract with Carlyle Systems the stipulation that a stress/capacity test (which we named "Phase X") be conducted. Because Carlyle had never installed a system as large as ours (over 100 terminals), the library administration felt the need to protect the university's investment by testing the proposed system in a standard way. Members of our implementation team had heard Salmon speak at a prior ALA meeting on the need for stress tests to safeguard customers in just such circumstances—a talk he may have regretted. Further, the university's computer group lent us a talented and supportive systems analyst who strongly recommended the stress/capacity test. We felt comfortable with the decision.

BOSTON UNIVERSITY'S PART IN THE STRESS TEST

Contractually, we asked that 100 terminals—live bodies, not a simulation—access our own database of 600,000 records using library-generated scripts over a three-hour period. We specified maximum transaction times, although those times were supplied by Carlyle; we were not trying to put an outrageous strain on the system, only to simulate what would happen once we had 100 terminals installed on campus. We also required a machine-generated hard-copy audit so that we could see later what had happened. Additionally, we specified that an impartial observer be present for the test. We attempted to assure the impartiality of the observer by splitting the fee equally with Carlyle. Finally, we agreed to pay for the cost of the test because such tests put enormous financial strain on fledgling companies.

SCRIPTS

Because we wanted to test our own database, we felt it necessary to design the queries ourselves. We formed a task force to go through the shelflist and compose scripts consisting of twenty searches each with a specific proportion of author, title, subject, and explicit Boolean searches. One month before the scheduled test we sent the scripts to Carlyle and to Epstein, our impartial observer, for approval.

FIRST STRESS/ CAPACITY TEST

Boston University sent two people from the Academic Computing Center, our systems analyst, and the library implementation team to the March 1986 Phase X; Epstein was also there.

Although tests were conducted on two separate days, we could in no way consider them successful. There were a variety of problems: inadequate air conditioning and power, some system-balancing errors, unruly student participants, and scripts that inadvertently set up destructive patterns.

We were enormously depressed, but fortunately Epstein was able to shed a perspective on what had happened during Phase X, show us the various options available to us, and help in rescheduling a second test. She brought objectivity to the failed first attempt and did much to facilitate the second test.

Julie Brown is Automation Specialist, Boston University Library.

SECOND TEST

On April 29, 1986, Carlyle conducted the second stress/capacity test. It was a stunning success. All the problems of the first test had been solved, and the system performed beyond expectations.

ADVICE

Based upon our experience at Boston University, we offer the following advice to others planning such a test.

• Consider how to get the necessary information in another way.

• If you need a stress test, make the provisions for it part of your contract.

• Try to defray vendor costs if possible.

• Know that you will be in a terrible dilemma if the stress test fails and you have based your entire contract on it. Try to look for positive indicators and keep your perspective.

• Let your staff know exactly what you are doing (we used newsletters). This not only informs staff but also keeps you honest. With so much riding on the outcome of the test, the temptation is to call a test a success when it is not.

• Consider hiring an impartial observer, particularly one who has had experience with stress tests. ■■

Measuring System Performance: The Vendor's Perspective

Stephen R. Salmon

This is the tale of a stress test. Most of the staff at Carlyle Systems were convinced by the time it was over that the phrase "stress test" applied to them, but it actually applied to a large, integrated system being installed by Carlyle at Boston University. As a part of the contract, Carlyle agreed to perform a benchmark, or performance, test prior to installation. The language of the contract, in describing its purpose, referred to a "stress test," and that's the phrase by which we came to know it, more intimately, as it turned out, than any of us intended. This paper describes

• what system we were testing,
• what we were measuring,
• how we measured it,
• what happened, and
• what conclusions can be drawn from the experience.

Carlyle had already installed a small system at Boston, and the stress test applied to a second, much larger system to be installed as the second phase of the contract. Actually, the system to be tested was not the system to be shipped and installed at Boston University but an even larger one, so that Boston could be assured that the system being shipped could be expanded in the future and still meet the required performance standards.

Specifically, the contract dictated a test of a system with 100 terminals accessing the full machine-readable bibliographic database at Boston University, estimated at about 600,000 bibliographic records. The test was to run a minimum of three hours with a volume of 1,500 to 2,000 transactions per hour. If the test were successful, it would demonstrate that the response time under these conditions would be less than six seconds on average and less than two seconds per 1,500 records found in Boolean searches. Carlyle also agreed to provide a log of all these transactions, with a variety of statistics and analyses.

The system to be tested included not only the 100 terminals and the 600,000 records already mentioned but also five disk drives to hold the database and all the indexes and ten different processors. The reason for ten different processors, for those unfamiliar with the Carlyle system architecture, is that we make extensive use of concepts called "distributed processing" and "parallel processing." In large library systems such as the one at Boston, with a number of branch locations, the user interface programs are run on what we call "remote

---

Stephen R. Salmon is Chairman, Carlyle Systems.

processors," located wherever there are clusters of terminals. They do all the terminal input/output functions, interpret the commands, and format the replies. They also contain all the error messages and the extensive array of help messages, so that these are instantly available, rather than having to be retrieved from a central source; and they act as intelligent multiplexers, concentrating signals from as many as twenty-four terminals to send to the central system.

At the other end, we use processors we call "database machines" to do all the database management functions—putting records on disk, managing the indexes, retrieving records from the disk, and so forth. In between are computers called "interface processors," which keep track of the state of everyone's transactions, do the Boolean combinations, pass messages back and forth, and monitor the system. This architecture gives us fast performance, because the different processors are all operating in parallel, and it gives us a lot of flexibility in constructing systems to meet various needs and sizes of libraries. For the Boston stress test, there were six remote processors, two database processors, and two interface processors, all running in parallel—the largest system we had built at that time.

We measured response time, transaction load, ability to handle a large database and many terminals, and reliability. The specifics are included below in "Boston University Benchmark Test: Facts and Figures" and "Carlyle Systems Performance Criteria," which contains the specific criteria we include in most of our contracts.

How we would actually run the test was not specified in the contract but evolved through negotiation between Boston and Carlyle. One key point is that Boston wanted people at the terminals. Carlyle had earlier developed a technique of capturing actual, real-life searches in transaction logs on an existing, installed system and then using these searches to test new systems. In this technique, the searches are fed to the system being tested by another computer at whatever speed is desired, the searches are executed, and the results are observed. The system being tested, of course, has no way of knowing whether the signals coming down the wire are generated by another computer or by a human pressing a keyboard; the performance results are the same. We had successfully used this method of testing to demonstrate performance and reliability in other contracts, notably for the large system we had installed at the New York Public Library, and we assumed initially that we would use it for the Boston test. For several reasons, however, Boston felt strongly that there should be people at the keyboards during the test, and that's what we did, with some interesting complications.

For one thing, this approach meant there had to be room for the 100 terminals and people to sit at them. The test took place in Berkeley, before we moved to our new, much larger building in Emeryville, so we had terminals strewn throughout our crowded quarters, in every conceivable place we could put a table or a desk: sixteen in the conference room, a few in this office, a few in that one, and wires strung everywhere.

It also meant that we had to find and hire the people. As long as we were going to use people, we thought we might as well make it as realistic as possible and use students, since they would be the principal users at Boston. This created a number of unexpected problems. First of all, we discovered (much to our surprise) that it is difficult to find 100 students who are willing to be paid for only a few hours work, even in a town with a large university and in a metropolitan area with a number of other educational institutions. Second, we learned that they may not show up; we were lucky if half of the ones we hired actually appeared within an hour of the time the test was supposed to start. Third, we quickly found, to our dismay, that college students don't always follow instructions. They become bored and start doing things their own way, regardless of how the test is supposed to be run. They want to read the help messages; they want to ask questions; they want to argue the merits of different system approaches, hardware, or software. We even discovered them reprogramming the function keys on some of the terminals!

In further pursuit of realism, we assumed that we would ask the students simply to do the kind of searches that students do. This became one of the most controversial issues between Carlyle and Boston, although I hasten to add that it was resolved amicably. The university library staff had prepared a set of several hundred searches, which we came to call "scripts," that they wanted used in the test. The searches were designed to insure that a variety of records, call numbers, and locations would be retrieved, and they used actual author, title, and subject headings from the Boston card catalog, so that the results would be known. The controversy arose because in our experience most users do not know and use library headings. Furthermore, the indexes chosen by Boston for its automated system were mostly keyword indexes, rather than heading indexes. This meant that almost every search would be a Boolean search, combining the words in each phrase, rather than a search on the heading as a whole, and we were concerned—unnecessarily, as it turned out—that the resulting response time would be unacceptable. The controversy was resolved when the library assured us that they "did not find it surprising, or unacceptable, that searches using those headings [took] longer to execute" and pointed out that the transaction logs indicated the system would perform such searches within the given time limits.

The actual test was run twice. To quote from the library's newsletter, "Carlyle's first attempts at running our Stress/Capacity test were inconclusive. That is to say, Carlyle's equipment did appear able to process the required number of transactions with our scripts searching our full database within the stipulated response times, but because of electrical and other technical problems outside of Carlyle's immediate control we never saw a full run of the 100 terminals for the requisite three hours." That's a fair and charitable summary: what the newsletter called "inconclusive" we tended to call "nightmarish."

First of all, the air conditioning broke down. I will leave it to the reader to imagine being in the same room with 100 students and 20 nervous staff members without air conditioning. Then we discovered that the night before there had been a power failure that damaged one of the disk drives. Then, midway through the test we began to smell something burning, and it turned out to be the main electrical transformer. Partly as a result of these problems, as the Boston newsletter said, we were unable to keep the system going long enough to complete a three-hour test.

In analyzing this first attempt during the following days, we discovered other problems.

1. In trying to add some features to the transaction log, we had introduced a bug; there was also a typo in some new software.

2. There was a hardware problem that took several days and several engineers to resolve. It turned out, to make a long story short, that our supplier had shipped us some processor boards that were different in a way they had neglected to tell us: they had a clock speed of 8Mhz instead of 6Mhz. When we changed the settings of the system to match the changed speed of the chips on the board, the problem disappeared, but it gave us fits during the test and for days afterwards.

3. We figured out why the system had been inexplicably slowing down as more and more searches were made. When we lost one disk drive because of the power failure on the day of the test, we managed to pack the full database and all the indexes onto the remaining four disk drives. It barely fit, and as a result there was very little room for transient data like the intermediate results of searches. As the day went on, the poor disk drives had more and more trouble finding work space to store data temporarily, and hence the system kept grinding slowly toward a halt. When we restored the missing drive, the problem disappeared.

We also discovered another interesting phenomenon involving the scripts. Apparently a pattern of commands had been developed by staff members at the library, then reproduced on each sheet of paper, and the content of the searches was filled in afterward. The searches were different, but the same pattern of commands and indexes was repeated on each operator's script, so that the same command functions and indexes were being impacted simultaneously by 100 terminals, in waves. This is roughly similar to sending troops marching in lock step across a bridge, which is a good way to have the bridge fall down.

After the first test, then, we made a number of changes. We found the bug I mentioned and corrected the board and the disk drive problems. We solved the problems created by the students simply by going to a regular employment agency. We mixed up the scripts to break the lock-step pattern and introduced more variety in the scripts. And thanks to some expert advice from consultant Susan Baerg Epstein, we made some changes in the *way* the searches were typed. Instead of having everybody start at once, we started them at intervals, with different kinds of searches and with different displays. On April 29, 1986, Epstein and the Boston contingent arrived again, and we ran the test a second time. This time we passed with flying colors, much to everyone's relief.

Looking back on the experience, we were able to identify a number of factors we believe are important to consider when planing such a test (see "Considerations in Running a Benchmark Test"), and I would like to stress one or two of them. First, make sure that you allow for a rerun, just in case you have power problems or other unpredictable circumstances; Boston University was clearly wise in this respect. There's too much at stake—for everybody—to chance it all on one roll of the dice.

Second, we were very fortunate in the attitudes and expertise of the people we were dealing with. The library staff were properly tough and exacting but at the same time very understanding and almost infinitely patient—even if at times a bit nervous. And we were fortunate that both Boston University and Carlyle Systems had agreed to have Epstein present as a consultant: she was impartial but helpful, and her extensive previous experience with such tests was very valuable.

And finally, there is one unshakable, unalterable, overriding, bottom-line conclusion we reached that was put most succinctly by one of our vice-presidents: we will never, ever, do this again!

BOSTON UNIVERSITY BENCHMARK TEST

Facts and Figures

On April 29, 1986, Carlyle Systems, Inc., conducted a benchmark test of the online catalog system scheduled for installation at the Boston University library. The test was conducted at Carlyle's home office in Berkeley, California, with representatives from the university's library and their academic computing and information systems offices in attendance. Library automation consultant Susan Baerg Epstein monitored the test for Boston University and Carlyle.

The contract specified the following test parameters:

- Approximately 600,000-record database
- 100 terminals in operation simultaneously
- Average volume of between 1,500 and 1,800 transactions per hour for three hours
- Average response of six seconds per single-word search
- Average response of one second per 750 records found for Boolean and phrase searches

The Carlyle system surpassed all contract specifications. Test parameters and results were as follows:

- Complete database totalling 588,706 bibliographic records was indexed according to Boston University's specifications
- 100 terminals were in operation simultaneously, entering a mixture of single-word and Boolean searches
- Volume averaged 2,025 transactions per hour
- Response time was less than one second per search for single-word searches
- Response time was 965 records found per second, on average, for Boolean and phrase searches

The first phase of the hardware installation took place in early June 1986 in the university's Mugar Memorial Library.

CONSIDERATIONS IN RUNNING A BENCHMARK TEST

1. Expense
   Cost of furniture and equipment required
   Cost of operators required
2. Realism
   Equipment to be tested
   Factors to be measured
   Type of searches to be used
   Timing of searches
   Patterns of searches
   Method of inputting searches
3. Timing
   Deadline dates
   Provision for reruns
4. Environmental requirements
   Power
   Temperature control
5. Expertise
   Vendor staff
   Customer staff
   Consultants
6. Attitudes
   Cooperation
   Patience and understanding
   Goals

CARLYLE SYSTEMS PERFORMANCE CRITERIA

A. Response Time

"Response time" is defined as the average length of time between the sending of a command to the system by pressing the return key on a terminal and the display of a prompting symbol or other indication that the system is ready to receive another command.

A "search" is defined as a command or statement typed on the keyboard of a terminal or read by a wand or laser reader.

An "explicit Boolean search" is defined as a search for two or more search terms or words combined with the words AND, OR, or NOT as logical connectors. An "implicit Boolean search" is defined as a search for two or more key words in the same index, not combined with the words AND, OR, or NOT as logical connectors.

With all terminals operational, the average response time for circulation charge and discharge shall be within two seconds 95 percent of the time and not exceeding five seconds; other searches shall be within five seconds 95 percent of the time and not exceeding ten seconds, except for Boolean searches, which shall not exceed one second per 750 records found under each term.

B. Reliability

"Downtime" is defined as the ratio of the total time the system or any equipment is unavailable for use during scheduled hours of operation, not counting time lost due to conditions beyond Carlyle's control (such as power failures or acts of God), to the total scheduled operating time, expressed as a percentage.

For the purpose of measuring downtime, the measuring period shall be a calendar month.

If the system is partially available but some items of equipment are not, the following table of coefficients shall apply in computing downtime.

Loss of access to the bibliographic database or patron database	1.00
Loss of access to a disk drive	1.00
Lack of availability of other items of equipment	1.00 divided by the number of such items

■■

Measuring System Performance: The Consultant's Perspective

Susan Baerg Epstein

For me, as a consultant, the Boston University stress test of Carlyle Systems was a very different kind of project because I was being paid by both parties; usually consultants work for the library. Additionally, this situation differed from those described by Clifford Lynch because we had added an extra level of stress by working within the context of a contract with extensive financial implications. The library is at great risk if it spends that money unwisely (i.e., the system dies six months after installation), and the vendor stands to lose a lot in both future sales and in actual dollars.

We have been calling it a "stress" test because that's what it is. You have also probably heard it called "benchmark" testing. In the Boston University (BU)-Carlyle contract it was called "Phase X" testing. The same Carlyle vice president who said "never ever again" also called this the "Endurance Test."

It is important to realize that any time you enter into one of these situations, both the library and the vendor should want it to succeed. If you do not want it to succeed, you should not be in the position of performing a test. It is simply too difficult to switch to another vendor. You may have lost your window on funding, and one of the things you certainly will have lost is the enthusiasm that your entire institution had for automation. You chose a vendor. In doing so, you decided that it was the best vendor possible for you, and you rejected the others. And, if the test fails, what do you do? Do you go back to your staff and tell them, "Well, we're going to take an also-ran?" This is a very important point: if this test fails in a big way, you are stuck with either taking another vendor or, even worse, giving up automation altogether.

Julie Brown and Stephen R. Salmon have stated that a stress test involves a great deal of work and that it should not be undertaken lightly. One of the BU staff members said, "In my experience, nothing is ever as easy as it seems, especially to the other interested party."

Susan Baerg Epstein serves as the president of her own consulting firm and has worked with many libraries as a consultant over the past nine years.

SCRIPTS

The scripts clearly were an issue of contention. But the issues were resolved. And in the BU-Carlyle situation, frankly, if the test had succeeded the first time, nobody would have ever looked at the scripts again.

There are some important factors in developing realistic scripts. The first is to look at what people really do, which is a major part of the contention over the Carlyle-BU scripts: the reality factor. And another consideration is, How often do they really do them? I have to admit to something I once did. With the first stress test I ran, I had a vendor say to me (fortunately in private and not in public), "I'd be very happy to run your test as soon as you can find an operator who can run three hundred PAC searches an hour on a terminal." We had calculated the number of searches we anticipated based on usage factors at the library, but we had neglected to divide that by the number of terminals available during that phase of the test.

I am currently working with a large library that is in its final acceptance test phase and has taken care of all the problems we've talked about. We used 3-by-5-inch cards. (That's what you do with old catalog cards when you get rid of the card catalog; you use them for testing.) On the backs of the cards we wrote "check out five books," "search on United States History," and so forth, and we basically built scripts from the cards. First we decided how many circulations we had to do, and we created enough cards to do five per checkout; we did this for all transaction types. We randomized the cards and sent them to the sixty locations. The first time we ran the test (and the BU-Carlyle test is not the only time the first test hasn't worked), it was an immediate disaster. In sending out the cards randomized, we had not counted on the fact that the library staff was going to alphabetize them! So, despite the fact that we knew that checkouts should not all be done in the first five minutes of an hour-long test, that was what was being done.

Scripts are difficult to develop. As Brown mentioned, one of the roles that I filled in this particular situation was not only that of a validator, but also that of an intermediary, somebody from outside. It's easy to get so wrapped up in the importance of what you are doing that you lose perspective.

BU-CARLYLE STRESS TESTS

The two tests themselves were interesting and, I think, touched on something Lynch mentioned when he talked about insight versus numbers. With every large test we have ever run, we knew within the first five minutes whether the three-hour test was going to pass or not. In the first test at Carlyle it was clear from the outset that it would fail, and gloom and doom did indeed prevail. The second test was so routine that it was downright boring. It was so straightforward that one of the staff members from Boston University and I, in the midst of this very important test, began to do some searches on our own that were not in the scripts. That's when we discovered that Boston University has over 69,000 occurrences of the word "United" in its database, and it appears over 68,000 times in combination with the word "States." If it does not appear with the word "States" it almost always appears with the word "Nations." (I don't know what you want to do with that piece of trivia, but we did find that out during the test.) It was so clear that the system would pass the second test that, even if something had gone wrong two hours into the test, I am certain that Boston University would have accepted the system.

RESPONSE TIME VERSUS THROUGHPUT

I have developed several conclusions over many painful days of stress tests.

I, too, prefer live operators. Lynch mentioned some of the difficulties in measuring response time, but when you are in a contract/profit situation, frankly, all those difficulties in measuring response time can also turn into ways that things can be manipulated. Lynch also asked, "What do you count when the system responds and says, 'I'm working on this?' " If you rely exclusively on response time as your measurement, you run the risk that your vendor will create what I call "gratuitous screens." Response time, in and of itself, doesn't tell you how many transactions you can get through the system. If a vendor puts in gratuitous screens, you can indeed get two-second response time, but it may take you ten screens to get what you want. Is that two-second response time better than a seven-second response time when you only have to use one screen? No, it is not. You need to balance response time with transaction mix. What can you put through the system? What can your users put through? That's a good reason for using live operators.

TEST SITE

It is almost impossible to do a live full-load test with more than one hundred terminals at any site other than your own. And one hundred terminals is not very many in today's automation environment. In doing a test at another vendor's site, we also had serious problems with the wiring hanging from the ceiling due to trying to fit a hundred terminals up and down the vendor's halls. We also have had staff problems. The test at Carlyle did not resort to what we used at another vendor's site where we hired friends and relatives of the company's employees.

Another point to consider is if the vendor has enough available space at an ongoing working operation, particularly if the system needs a computer room, to hold a CPU and disk drives sufficient for your system. This is very unlikely. And, if you are trying to conduct this test prior to your signing a contract, you've put the vendor at tremendous financial risk. One of the vendors I worked with later told me that they figured it cost them $60,000 to run a test at their site that was successful on the first try.

CONSIDER OTHER TESTS

There have been a number of stress tests, and if you are looking at doing testing, you ought to consider the results of some of them: the Data Phase test at Orange

County, a Dynix Test at Dynix for Ramsey County, a Geac test for Incolsa, an Eyring test for Pikes Peak, and a DRA test for Los Angeles County Public Library that was done in simulation.

Why does a vendor ever go through this? I think many of us can see why a library wants to go through it. The library wants to have absolute positive proof that the system will run. But why does the vendor go through it? In my experience every time this has occurred, it has been because the vendor is attempting a significant upscaling of the system configuration offered. They are fully confident that they can manage this, but they need to prove it. And they want the business that is represented by the stress test very badly. The library should decide whether or not its situation warrants a stress test before ever suggesting one. These tests are not easy, and they are very expensive. Not every user should do one, but almost every vendor is going to have to do one sooner or later unless they are very lucky. One of the reasons to undertake a stress test is, frankly, not necessarily that vendor's fault. We have had significant failures in this marketplace: many underconfigured systems, systems that never did perform, systems that were not well integrated, systems that could not work with any known telecommunications configuration, systems that worked in simulation but not in reality.

Once a vendor has done this type of testing, they probably will not want to do it again unless the vendor attempts a significant upscaling. This would have to be a significant upscaling—not over what they had previously tested but over what they have installed and operated.

With all of these tests that have now worked in situations with large transaction loads and over one hundred terminals, I suspect that you are not going to see any vendors who have done a test doing another stress test until systems get into the three-, four-, or five-hundred-terminal range. Even then I have some doubts as to whether they'll do live testing or simulation.

If your library is not pushing the boundaries of a vendor's performance and pushing well beyond, you should be able to use the results of these previous tests to get some indication of how well the system can perform for you. I see two possible barriers to expanding the system if it is underconfigured for your performance needs, two barriers to making the system do what you need it to do. One barrier is money and the other is technology. Libraries tend to worry most about money, but, frankly, the biggest concern is technology. If your vendor's system has a limit it cannot surpass, it doesn't matter how much money you have; the system can't work. On the other hand, contracts provide a number of ways to protect against underconfigured systems that will mitigate the money issues. However, if you seek these protections you cannot at the same time get the rock-bottom price. But those are issues you have to balance.

One of the tendencies I am seeing is that when we have publicity for such issues as the Boston University-Carlyle stress test or some of the vendor failures, we see a change in library actions. Libraries that are installing "vanilla" systems, systems that a particular vendor installs regularly, are asking for extraordinary contract protection clauses. I think these libraries are hurting themselves. Although I am not advocating that everyone perform this kind of stress test, everyone should do some sort of final acceptance test. Those tests take place on your system in your library. That eliminates some environmental and operator problems, but it doesn't eliminate your responsibility for doing rigorous testing with good scripts. I am disappointed when I work with a library and struggle in contract negotiations to get provisions for a good test, but then the library doesn't do the test and later has capacity problems.

NEED FOR STANDARDS

Looking at standards, what comes next, and what do we need to do? One of the things I would like to be able to do is to use these test results with some level of confidence, both for myself and for my clients. Today we have discussed one test. Each test of a commercial vendor's system has, so far, been done at the instigation of an individual library with that library's individual specifications. That library paid for the test. Whether they paid for it directly as Boston University did or whether the payment was buried in a bid cost, they paid for the test. So the library really didn't have much interest in designing a universally applicable test. It's a question of research results versus information needed to manage the library.

I am curious as to whether it is possible for us to develop some comparatively standardized test that can at least be used as a basis for deciding whether a system has sufficient capacity. There may, however, be so many combinations that this is impossible. Even trying to balance circulation loads and PAC loads may make this impossible.

These tests aren't easy. They certainly aren't fun, except when they succeed easily. They are an important step along the way toward the future of library automation, but do consider carefully your real needs before you enter into a stress test. ■■

A Framework for the Development of Performance Measurement Standards

Paul Evan Peters

The following insight to performance testing is owed to a friend of mine who is an aeronautical engineer. Called the "chicken" testing of new jet engine designs, it represents what I regard to be the ultimate example of "stress" testing, a topic of current great concern. Let me tell you what a chicken test is. First, you mount in a horizontal position a newly designed and fabricated jet engine that's ready for testing. Second, you position a pneumatic cannon directly opposite the air intakes of the jet engine. Next, and here is where it starts to get interesting, you fill the barrel of the pneumatic cannon with dead chickens. If you are still with me, I am sure you have already figured out the next steps: you fire the jet engine up to its maximum power and use the pneumatic cannon to blow the dead chickens into the engine's air intakes.

If the truth be known I would have begun with this story regardless of whether it had a lesson to offer for the matters at hand. At the very least, the story conveys a powerful visual image that approximates the feelings reported by many of those brave souls who have been through a carefully designed and thoroughly pursued stress test of a library automation system. However, I believe the chicken test illustrates two important principles that should provide the same sort of foundation for the performance testing of library automation systems that they do for the performance testing of new jet engine designs. The first is that the most difficult performance tests should be done first. The simple wisdom of this principle is that it is much more difficult and expensive to change a design to measure up to extreme performance requirements at an advanced stage than it is to change one that is at an early stage. The second principle is that *how* a system fails under pressure is as important, if not more important, than *whether* it fails. Few new jet engine designs pass the chicken test by continuing to operate. It is the engines that fail without blowing up and, by implication, taking a wing and probably an aircraft with them that are said to "pass." The testing of library automation systems should be similarly concerned with the effect of system failures on the entire array of activities the system has been implemented to support.

My point of view is that of the "standards world," and I question whether performance measurement standards are possible or timely. I am going to assume that the answer to this question is *yes*. I hope to give a feel for what structure such standards may eventually take and what means will be required to produce them.

Clifford Lynch gave us a good working definition of what a standard is when he remarked that much progress could be made if we would all just accept a certain number

Paul Evan Peters is Systems Coordinator, New York Public Library.

of arbitrary definitions. Consensual, arbitrary definitions establish a shared baseline against which things can be compared, and this is exactly what we want from performance measurement standards. This is one of the reasons why I believe that developing such definitions for "response time" and as many as possible of the other measures, regardless of the complexities, would be tremendously useful. I also believe that doing so is an idea whose time has come.

I am going to focus on technical standards and not spend any time on contractual provisions, recommended performance levels, and the many other things that the word "standard" may bring to mind. These are important and interesting, but they turn in the main on the codification and regulation of matters that are not first and foremost technical in nature.

My remarks are based on information and opinion I have gleaned from a number of sources, but two have been especially helpful. The Association for Computing Machinery has a special interest group on performance measurement, and it publishes an incredibly useful quarterly newsletter called *Performance Evaluation Review*. If you like mathematical models and the definitions on which they are based, this is an excellent place to look. I would also like to call attention to four standards that have been produced by Accredited Standards Committee X3 (Information Processing System): X3.44–1974, entitled *Determination of the Performance of Data Communications Systems;* X3.79–1981, entitled *Determination of Performance of Data Communications Systems that Use Bit Oriented Control Procedures;* X3.102–1983, entitled *Data Communication User Oriented Performance Parameters;* and X3.141–1987, entitled *Data Communications Systems and Services: Measurement Methods for User Oriented Performance Evaluation*. These are, in my estimation, the four best performance measurement standards in the field of information processing.

It is important to recognize that performance measurement is a special case of performance evaluation and that performance evaluation is a special case of system evaluation. Also by way of context for performance measurement, it is important to recognize that performance evaluation is but one of the three major concerns of system evaluation. The first is functional evaluation, which is concerned with whether an individual system has the features you need. The second is economic evaluation, which is concerned with whether you can afford a given system and whether its benefits are worth its cost. Our present interest is the third major area of system evaluation—performance evaluation—and it is concerned with whether a given system has sufficient capacity to meet the demands that will be placed on it.

A current trend in the library automation system marketplace is that performance evaluation is becoming more and more of a concern as a result of the growing maturity of that marketplace. In a new technological marketplace, functional evaluation is the dominant mode of system evaluation because the overriding concern is one of simply finding a system that offers the functions that are needed. Since there is no alternative to such a system once it is found, it is purchased if it can be afforded. Sooner or later a marketplace matures to the point at which buyers have to make choices between systems that appear to be roughly functionally equivalent and equally affordable. This situation gives birth to the concern for performance evaluation. I believe that the library automation system marketplace has reached this point of maturity, and we all are going to be spending more and more time talking about performance evaluation because it, rather than economics or functionality, is going to be the major domain of competition and capability.

We are interested in performance evaluation because we are concerned about the long-term price/performance profile of the systems we are buying. First, we are interested in assessing whether the systems we buy meet the performance requirements expressed in our contracts and regulated through our acceptance tests and system management instruments. But we are also interested in upgrading our systems. Let's face it, we all want to add more terminals, get gifts from heaven that allow our recon programs to exceed their most optimistic projections, and otherwise receive good news that places new demands on our library automation systems. We need to know precisely whether our existing systems can be expanded to handle the load or whether new systems will be needed. We also need to learn answers to such questions in an orderly fashion.

Performance measurement is a subspecialty of performance evaluation. I will focus on the quantitative aspects of performance measurement. There is more to performance evaluation than measuring things, but, for me, performance evaluation begins with measuring the things that can be measured. I do not mean to suggest that quantitative measures exhaust the set of performance measures that can be dealt with in a valid and reliable fashion. Rather, I see performance measurement as the quantitative expression of the desire for performance evaluation.

One of the ironies we face is that just as we have come to recognize all the good reasons we have for doing performance measurement, faith is being lost in the traditional measures by which performance is assessed. For instance, "MIPS" (Millions of Instructions per Second), as I believe Stephen R. Salmon noted, cannot be counted upon as a performance measure of parallel or pipeline processors as it was counted upon as a performance measure of the "von Neumann" computing machines that got us to where we are today. In addition, as Lynch has helped us to understand, "response time" is a very imprecise measure and perhaps was not a good measure in concept to begin with.

I accept everything covered here about "response time." I would like to add that response time is a failure measure in that it is much easier to say when response time is poor than when it is good. Furthermore, I am not convinced that the difference between a one-second response time and a two-second response time is that a one-second response time is twice as good as a two-second response time. Is it, or isn't it? I submit that a two-second response time is just about the same as a one-second one. It certainly is not twice as bad. But, I also submit that a ten-second response time is much worse than twice as bad as a five-second response time. The point of all this is that response time is much too elastic a measure on which to base the sorts of costly and important systems decisions we have to make. It is this elasticity that signals the degree to which it depends upon human factors at least as much as it depends upon system ones. It is difficult and perhaps even unwise to use response time as a comparative performance measure when it is known to vary so strongly as a function of end-users' characteristics.

I wish now to follow up on two words that are common in the literature of performance measurement: parameters and factors. A "parameter" is an output performance measure, whereas a "factor" is an input performance measure. A performance parameter is what is known in statistics as a "dependent" variable. It is what you care about most and can experience, and it is what the configuration of the system can be adjusted to affect. This contrasts with a performance factor, which is known in statistics as an "independent," or "controlling," variable. For example, "response time" and "mean time between failure" are examples of performance parameters, whereas the baud rate by which one communicates with a system, the size of a system's database and its median record, and the rate at which transactions arrive are all examples of performance factors.

It is also important to keep in mind that there is more than one type of performance measurement. First, there is "absolute" performance measurement, which endeavors to estimate performance parameters and can be illustrated by the attempt to estimate a system's response time to be "X plus or minus Y with a confidence of Z." Absolute performance measurement is used to determine whether a system meets the performance requirements that have been set for it. There is also "comparative" performance measurement, which tries to quantify the differences, if any, between two systems. Comparative performance measurement is used to select the best of competing systems that are functionally and economically equivalent and to plan and gauge system upgrade programs. Finally, there is "diagnostic" performance measurement, which strives to decompose a performance parameter and to explain it in terms of its component performance factors. Diagnostic performance measurement is used to identify the performance factors with the greatest influence on a given performance parameter, thereby identifying those aspects of a system's configuration that should be changed first to improve the performance parameter of interest. Steve Salmon called our attention to this when he recounted the performance problems that resulted from putting too much data on a single set of disks.

The last thought I would like to introduce before turning to performance measurement standards per se is an echo of something both Julie Brown and Susan Baerg Epstein stated: it is extremely important to make sure that performance measurement exists in a general problem-solving framework. The objectives of performance measurement must be kept in mind. You want accurate results. You want results that are applicable to actual decisions you are trying to make, rather than to all conceivable decisions you may want to make. And, you want results that are affordable. This last objective deserves special mention. We must never lose sight of the fact that our ultimate objective is improved service to and systems support of our clientele and staff and not the construction of an exact science of performance measurement. This objective rules in favor of doing whatever performance measurement we can when we can and not letting statistical conventions, real or imagined, constrain our actions more than they facilitate them.

The time has come to zoom in on the standards and what needs to be done to start constructing standards in this area. The first thing we need is a "reference model," and in our case we may need more than one. A "reference model" (a term popularized by the "open systems interconnection reference model" that has been so instrumental to the development of the suite of standards attributable to the Linked Systems Project) is, first of all, a standard that establishes a generic system description that distinguishes between "actors" and "resources" fundamental to a given process. Reference models also specify the characteristics of the "interfaces" and "events" by which actors interact with each other and with the resources available to them. The reference events codified by an individual reference model can be regarded as the "primitive transactions" that define the behavior of the system in question. I am not optimistic that a single reference model for library automation systems would be possible or advisable. It seems more promising to begin on a system-by-system basis (e.g., circulation, acquisitions, catalog maintenance, etc.), letting differences and similarities reveal themselves rather than forcing a common framework on everything from the outset.

Once we have a reference model (or a family of reference models) in hand we can turn attention to a second level of performance measurement standards: standard "units and measures." In this context the term "measures" is synonymous with performance parameters and factors as defined earlier, and the term "units" refers to how we decide to quantify the measures in question (e.g., bits per second, characters per record, etc.). The reference model or models together with the units and measures provide a complete conceptualization of the performance of a given system. They define the system in question, its basic transactions, the parameters by which the performance of the system is observed, the factors that determined the values assumed by the parameters, and the units by which the parameters and factors are quantified.

There is a third and final level of performance measurement standardization. At this level can be found standards addressing data collection, analysis, and presentation issues as they pertain to the measurement of performance parameters and factors. Sampling frames and rates are a concern, as are the appropriate uses of means, medians and percentiles, variances and standard deviations, and other summary descriptive statistics and tables, graphs, bar and pie charts, and other means of visual display. The most sophisticated standards at this level codify complete test programs, such as sets of search strategies, mixes of transactions, and the like. Such test programs are exemplified by the "sieve," "whetstone," "core seek," and "Norton computing index" tests commonly used to rate the performance of personal computers. Performance measurement standards at this level are possible only at the most advanced stages of development and only after years of both conceptual and practical efforts. They represent a particularly compelling goal; the sooner we start on this long road the better.

In conclusion, I believe that the National Information Standards Organization (NISO) is the appropriate standards development organization to sponsor performance measurement standards for library automation systems. The only real conceivable alternative is Accredited Standards Committee X3 (Information Processing Systems), the standards development organization whose data communications performance measurement standards were mentioned above. But X3's focus is on basic information systems and technologies, whereas NISO's is on the application of such systems and technologies to the problems of publishing, libraries, and information services.

I suspect that NISO will begin this work on its own initiative by forming a standards committee and charging it to attempt some first step in this area, the one recommended above or some other. This is not the only way NISO forms and charges its standards committees. A frequent alternative is exemplified by the Standards Committee EE, which developed the CD-ROM volume structure and file format standard Z39.60. This standards committee was formed and charged to process a draft standard that had been produced under the auspices of an ad hoc task group. Regardless, once a draft standard has been prepared by a NISO standards committee and approved by the NISO voting members it is forwarded to the Board of Standards Review of the American National Standards Institute, where it is registered as an American National Standard. The next step is to submit a standard to Technical Committee TC46 (Documentation) of the International Standards Organization for processing as an international standard.

I would like to thank the organizers of the TESLA program on performance measures for the opportunity to contribute and to benefit from the contributions of my colleagues. I hope this article has presented an array of background information and made a variety of useful distinctions. I further hope I have provided ample reason to believe that some day in the not-too-distant future we may all enjoy the availability of a robust set of library automation system performance measures. ■■

Twenty Years Ago in JOLA

To the question, "Would it be better to return to a completely manual system for ordering books?" the answer by the Order Section has always been "No, retention of the automated system for fiscal control and voucher preparation is preferable, even with the patched system at hand." Nor should it be forgotten that the book order system as originally designed worked well until the demand on it exceeded its production capacity. Also to be recognized is the gain in experience and insight by the library staff during these three years. Reading about or visiting someone else's work is enlightening but day-to-day work brings an understanding for which it is difficult to obtain a substitute.

Lawrence Auld, "Automated Book Order and Circulation Control Procedures at the Oakland University Library," *JOLA* 1, no.2:108 (June 1968).

Communications

Sheet Music Index at the University of Colorado: II

Nancy F. Carter

Since the sequel is such a popular concept these days, why not a sequel to an article from *Information Technology and Libraries*? The article of interest, appearing on p.52–55 of the March 1983 issue, is "Sheet Music Index on a Microcomputer." In the intervening 4.5 years this index has remained alive and well and has evolved. I am happy to report that it has undergone many changes, all of them beneficial.

RECAP: MICRO

In 1982 the Music Library at the University of Colorado in Boulder owned approximately 2,000 scattered pieces of unused vocal sheet music resulting from many years of gifts and neglect. The decision was made to index this collection with four points of access: title, composer, lyricist, and first line. Because the main library owned a Radio Shack model II microcomputer, it seemed that a cheap and convenient way to make our indexes was to use it with an off-the-shelf program called Profile. A disadvantage of this plan was lack of space on the computer disks. Since each disk held 800–1,000 records, we had no choice but to use multiple disks. We decided on six disks, using division by date of publication in order to leave room for future expansion. Our divisions were

Disk 1	1800s and No Dates
Disk 2	1900–1909
Disk 3	1910–19
Disk 4	1920–29
Disk 5	1930–49
Disk 6	1950–

Nancy F. Carter is Head, Music Cataloging Unit, Music Library, University of Colorado, Boulder.

Thus, when searching for a title one had to search in six alphabetical lists rather than one, but we could live with that. It was better than no index, which was what we had before.

The indexes proved very useful, and use of the material was brisk from the start. Our plan was to produce an updated printout every summer, and that is exactly what happened. It was gratifying to receive so many donations of sheet music that an annual update was necessary.

SECOND STAGE: UNIVERSITY COMPUTER

We functioned quite happily in this manner until the update of summer 1985 was concluded, when we were forced to acknowledge that our disks (all six of them) were full and could hold no more. By the time our 1986 update rolled around we would need to add new disks or transfer our data to a larger machine. The former option was very undesirable, so we started negotiations with our university computing center. The advantages of using the university computer were obvious: unlimited expansion, plus indexes in one alphabetical list instead of six. The main difficulty lay in acquiring a program that would allow the micro to talk to the VAX computer. The head of library computer services converted the data on Profile to ASCII format using Scripsit. He then sent the data to academic computing services, where one of their technicians modified and formatted the data and created an index generator so that the information would be sorted and printed. After admirable work by both parties, we received in August 1986 a very handsome printout consisting of five indexes sorted by title, first line, composer, lyricist, and year—we had made progress and were proud of our product. At this time, our database consisted of 8,000 titles,

TITLE	COMPOSER	LYRICIST	FIRST LINE	YEAR	COMMENTS
LOVE'S OLD SWEET SONG	MCLLOY, J.L.	ANON.	ONCE IN THE DEAR DEAD DAYS	1935	
LOVELY TO LOOK AT	KERN, JEROME	FIELDS	CLOTHES MUST PLAY A PART TC	1935	
LULLABY OF BROADWAY	WARRN, HARRY	DUBIN	COME ON ALONG AND LISTEN TO	1935	
LULLABYE OF BROADWAY	WARREN, HARRY	DUBIN	COME ON ALONG AND LISTEN T	1935	
LULU'S BACK IN TOWN	WARREN, HARRY	DUBIN	GOTTA GET MY OLD TUXEDO	1935	
MAN ON THE FLYING TRAPEZE	MANOLOFF, NICK	MANOLOFF	ONCE I WAS HAPPY BUT NOW I'M	1935	
MAN ON THE FLYING TRAPEZE	TRAD.,	TRAD.	HE'D FLOAT THROUGH THE AIR	1935	
MAY I SLEEP IN YOUR BARN TONIG	MANOLOFF, NICK	Anon.	CAN I SLEEP IN YOUR BARN TONIG	1935	
MAYBE	MADDEN, FRANK	FLYNN	YOU'VE COME TO TELL ME THAT	1935	
MELODY IN F	RUBINSTEIN, ANTON	CASTILLO	SWEET MELODY SEEMS TO FLOAT	1935	
MOON COUNTRY	CARMICHAEL, HOAGY	MERCER	I LONG FOR THAT MOON COUNTRY	1935	
MOON OVER MIAMI	BURKE, JOE	LESLIE	TROPICAL TWILIGHT DESCENDING	1935	
MOST BEAUTIFUL GIRL IN THE	RODGERS, RICHARD	HART	WE USED TO SPEND THE SPRING	1935	
MUSIC GOES 'ROUND AND 'ROUND	FARLEY, EDWARD	HODGSON	ONE NIGHT WHILE PLAYING IN THE	1935	
MY EXTRAORDINARY GAL	SHAND, TERRY	SHAND	IT'S NOT AN ORDINARY GAL THAT	1935	
MY SUNNY HOME IN SUNNY TENN	CALLJU, JOE	CALLJU	I WANNA BE BACK WHERE THE M	1935	
NEARER MY GOD TO THEE	DIBBLE, HORACE	DIBBLE	NEAR MY GOD TO THEE NEAR TO TH	1935	
NIGHT IS YOUNG	ROMBERG, SIGMUND	HAMMERSTEI	FAILING LIGHT OF FALLING S	1935	
NOBODY KNOWS THE TROUBLE	TRAD,	TRAD.	NOBODY KNOWS THE TROUBLE I'V	1935	
O HOLY NIGHT	ADAM, ADOLPH	ADAM	O HOLY NIGHT THE STARS	1935	*CHRISTMAS
O SOLE MIO	CAPO, EDVARDO	CASTILLO	HOW MANY SUNSHINES HAVE I	1935	
ON TREASURE ISLAND	BURKE, JOE	LESLIE	I SAILED AWAY TO TREASURE	1935	
OREGON TRAIL	DE ROSE, PETER	HILL	AN OPEN RANGE AHEAD A BLANKET	1935	
ORPHEUS' SELF MAY HEAVE HIS	HANDEL, GEORGE F.	HANDEL	ORPHEUS' SELF MAY HEAVE HIS	1935	
PAGE MISS GLORY	WARREN, HARRY	DUBIN	WHEN THE MOON IS NEW	1935	
PICTURE ME WITHOUT YOU	PORTER, COLE	PORTER	PICTURE HENRI FORD WITHOUT	1935	
PLEASE BELIEVE ME	JACOBS, AL	YOELL	HOW COULD I EVER DREAM THAT	1935	
RED RIVER VALLEY	MANOLOFF, NICK	MANOLOFF	FROM THE VALLEY THEY TELL	1935	
RED SAILS IN THE SUNSET	WILLIAMS, HUGH	KENNEDY	TWAS DOWN WHERE FISHER FOLK	1935	
RHYTHM OF THE RAIN	STERN, JACK	MESKILL	I LOVE THE RHYTHM OF THE RAIN	1935	
ROBIN SANG IN THE ELM-WOOD TRE	KOUNTZ, RICHARD	WICKOFF	IT WAS JUST AS STARLIGHT TWINK	1935	
ROLL ALONG PRAIRIE MOON	FIORITO, TED	TILZER	THERE'S A WONDERFUL LIGHT I	1935	
ROSE IN HER HAIR	WARREN, HARRY	DUBIN	LOVE IS A ROSE AND IT GROWS	1935	
SERENADE	SCHUBERT, FRANZ	CASTILLO	NIGHT WAS YOUNG WHEN YOU C	1935	
SHE SHALL HAVE MUSIC	MURRAY, ALAN	BRANDON	SHE'LL PASS THROUGH THE WOODLA	1935	
SHE SHALL HAVE MUSIC	HOFFMAN, AL	SIGLER	SHE SHALL HAVE MUSIC WHERE	1935	
SILVER THREADS AMONG THE GO	DANKS, H.P.	BEXFORD	DARLING I AM GROWING OLD	1935	
SONG OF THE OPEN ROAD	MALOTTE, ALBERTHAY	MALOTTE	WHAT IN THE WORLD COULD BE SO	1935	
SONGS MY MOTHER TAUGHT ME	DVORAK, ANTON	DVORAK	SONGS MY MOTHER TAUGHT ME	1935	
SORROW OF SORROWS	LENORMAND, RENE	ROSS	O STAR OF HEAVEN O SAINT OF MY	1935	+IN FRE
SOUVENIR	DRDLA, FRANZ	CASTILLO	THE DAY YOU WENT AWAY BLUE	1935	
SPRING SONG	MENDELSSOHN, FELIX	CASTILLO	SPRING BEAUTIFUL SPRING HAS	1935	
ST. LOUIS BLUES	HANDY, W.C.	HANDY	I HATE TO SEE DE EVENIN' SUN	1935	
STAIRWAY TO THE STARS	MALNECK, MATT	PARISH	LET'S BUILD A STAIRWAY TO	1935	
STILL AS THE NIGHT	BOHM, CARL	CASTILLO	STILL AS THE NIGHT DEEP AS	1935	
STRANGER OF GALILEE	MORRIS, MRS. C.H.	MORRIS	IN FANCY I STOOD BY THE SHO	1935	
SUMMER TIME	GERSHWIN, GEORGE	HEYWARD	SUMMER TIME AN' THE LIVIN' IS	1935	
TAKE ME BACK TO MY BOOTS AND S	SAMUELS, WALTER	POWELL	TAKE ME BACK TO MY BOOTS AND S	1935	

Fig. 1. Music Library's sheet music index by year and title.

a 400 percent increase over a four-year period.

Now our yearly updates would consist of keying data into the micro, then transferring it to the VAX for format and printout.

FINAL STAGE: ONLINE

In the meantime the university libraries changed from cards to an online catalog, closing the card catalog in September 1986. At present the libraries' catalog is 80 percent online. Our system is the Colorado Alliance of Research Libraries (CARL), a consortium made up of five university libraries and one public library; CARL has also acquired some associate members, the functions and privileges of which vary. Its Public Access Catalog (PAC) offers a word search (titles, subjects, series, and contents); a name search (authors, performers, added entries); and more specific searches by call number, series, and exact title.

As our clientele became more accustomed to the online catalog, we began to realize that sometimes a patron would type a song title into PAC and, after failing to get a hit, would leave the library without asking for assistance. That, of course, had always been a problem. The sheet music index had been available only by request. While we have tried to advertise its existence, it is not possible to reach everyone or persuade them to ask for assistance. Thus we began to consider the possibility of putting our index online so that it could be searched along with all other bibliographic records in our libraries. A CARL representative assured me that the technology would create no difficulties since all the data were already on tape. Each entry needed to be fitted into a "mock MARC" format, which we did in just a few minutes. A sheet from one of the printed indexes is shown in figure 1. The MARC format's field numbers were assigned as follows:

Title	245	First line	740
Composer	100	Year	260
Lyricist	700	Comments	500

We needed call numbers for our entries so that the data would remain together in one area of the libraries' database. We decided on MUSICPOP because there are no LC call numbers between *MT* and *N*. It was decided later to attach the year of publication to the call number. Now the entire index can be viewed on "call number browse": anyone who wishes to see all pieces in the collection published in 1935 would enter "MUSICPOP 1935." Having the year attached to the call number is also helpful because the music is boxed by year of publication. Figure 2 shows one of the screens obtained while browsing on PAC by call number with publication year 1935. Figure 3 shows a full record of one entry. Print

```
22 Pestalozza a                                CU   MUSIC  SPEC    1935
     Ciribiribin                               MUSICPOP 1935

23 Manoloff nick                               CU   MUSIC  SPEC    1935
     Climbing up the golden stair              MUSICPOP 1935

24 Donaldson walter                            CU   MUSIC  SPEC    1935
     Clouds                                    MUSICPOP 1935

25 Weldon frank                                CU   MUSIC  SPEC    1935
     Dancing on a roof top                     MUSICPOP 1935

26 Bernard felix                               CU   MUSIC  SPEC    1935
     Dardanella                                MUSICPOP 1935

27 Sosenko anna                                CU   MUSIC  SPEC    1935
     Darling, je vous aime beauco              MUSICPOP 1935

28 Manoloff nick                               CU   MUSIC  SPEC    1935
     East bound train                          MUSICPOP 1935

<RETURN> TO CONTINUE DISPLAY
ENTER <LINE NUMBER> TO DISPLAY FULL RECORD
<P>REVIOUS FOR PREVIOUS PAGE OR <Q>UIT FOR NEW SEARCH
WORKING
```

Fig. 2. MUSICPOP screen showing some selections from 1935.

```
AUTHOR(s):       Manoloff, Nick
TITLE(s):        Climbing up the golden stair

                 1935

OTHER ENTRIES:   Lyricist: Manoloff
                 First line: Then hear them bells a ring

CALL #: MUSICPOP 1935          LIBRARY:        MUSIC  SPEC

<RETURN> to continue, <Q>UIT for a new search, or <R> to REPEAT this display
<S> TO SELECT THIS RECORD:
```

Fig. 3. MUSICPOP's full record of one entry.

constants were supplied by CARL in the 700 and 740 fields.

Adding the sheet music index to our online catalog has proven successful. The word search or exact-title search on PAC will now bring up records for sheet music as well as other materials. A name search for Irving Berlin, for example, will now show that we have many of his songs in sheet music form. In the case of some of his earlier pieces, these sheets represent our only copies. Even if only a few words from the first line are remembered, the piece can still be found through the keyword search. To use the printed index, the patron had to know the first word of the first line or title.

Another advantage of the online index is easier maintenance. Corrections and new entries are entered on the bibliographic maintenance program—no more waiting until next summer—no more work sheets at a remote terminal for entry purposes. Now the data can be entered directly from the sheets of music in our library. This alone will save us much time and error. In addition, we have not lost the advantages of a printed index, which CARL has agreed to produce for us once a year.

Our index has come a long way. Old popular sheet music has become a legitimate study in music, and our index will help. Now there is nothing left to do except maintain it. However, if we had subject access only . . . . ■■

Twenty Years Ago in JOLA

The essential questions therefore became: Off-line or on-line access? Batch or real-time processing? . . .

The feasibility of on-line access has been widely demonstrated in the research and business world. Remote, on-line computer processing is clearly a common course of the near future. Equally predictably, it will steadily give more favorable cost/value ratios as machine costs decrease and labor costs mount. In sum, the Technical Information Libraries concluded that an on-line system was worth the price. Only an on-line approach would meet the overall objectives for a new system and offer advantages sufficient to justify conversion effort at this time.

R. A. Kennedy, "Bell Laboratories' Library Real-Time Loan System, (BELLREL)," *JOLA* 1, no.2:131–32 (June 1968).

News and Announcements

LITA Announces National Conference Program

Technologies that make a difference in library and information services will be the focus of the second national conference of the Library and Information Technology Association in Boston on October 2–6, 1988. The theme is "Effective Technology, Excellent Service: Putting the Pieces Together."

Conference participants, an expected 2,000 librarians and information science professionals, can choose from a variety of topics. On Sunday, October 2, six ninety-minute technology seminars will introduce and provide an overview of expert systems, microcomputers, optical information systems, desktop database building and publishing, telecommunications, and integrated library information systems.

On Monday through Wednesday there will be forty ninety-minute technical sessions, combining papers accepted as a result of the conference call for participation, invited papers, speakers, and panelists.

Technical session topics include microcomputer applications, preservation technology and disaster planning, vendor viability, authority work in an online environment, retrospective conversion, extended work in an online environment, retrospective conversion, extended uses of online catalog data, relations between libraries and computer centers, and applications of telefacsimile.

On Monday through Wednesday conference participants can also attend new product reviews, in-depth discussions, and demonstrations of newly introduced or enhanced products and services offered by selected exhibitors; or attend showcases featuring real-life uses of new technologies as librarians and information science practitioners share their most recent developments and projects. Exhibits from more than 150 companies will demonstrate the latest library and information technology products and services.

Visits to some of the outstanding libraries in the Boston area will give conference attendees an opportunity to see a broad spectrum of automated library systems in full operation.

Sunday and Tuesday schedules will provide time for relaxed conversation with other conference participants, either at the opening conference reception on Sunday evening at the Boston Public Library or at the gala clambake on Tuesday evening at the Sheraton Boston Hotel and Towers.

On Thursday, following the conference program, there will be six all-day workshops requiring separate registration. The workshops are

"Design and Use of Online Authority Control Systems"
"Hacking IBM PC System Hardware and Software Internals"
"The Apple Macintosh as a Library Workstation"
"Preparing for Serials Automation"
"Library Systems Facilities Management"
"Retrospective Conversion and Other Data Preparation, Profiling, and Loading Activities"

For further information, write or call LITA, 50 E. Huron St., Chicago, IL 60611-2729; (312) 944-6780. ■■

University of New Mexico Signs with Carlyle

The University of New Mexico has signed a contract with Carlyle Systems, Inc., for purchase of the Carlyle automated library system, which will be installed in six phases. The initial system includes the online catalog, cataloging input/edit, and OCLC interface. Circulation, authority control, and an interface to the INNOVACQ acquisitions system will be added in later phases of the installation; it will eventually support 250 terminals and provide access to more than two million records. The database will represent the holdings of several UNM libraries, including the general, law, and medical libraries.

The system has already been successfully linked to the university's Ungermann-Bass local area network. Besides the Carlyle test database, the network provides dial-up access to the libraries' LS/2 (DataPhase) and INNOVACQ systems. It is the intention of the UNM General Library to offer a variety of databases in addition to bibliographic records. ■■

CompuSystem II for Binding Preparation

CompuSystem II is a new software package written and developed by the information services department of The American Companies, Inc., according to Carolyn Cobb, corporate IS manager. A standalone, automated, data retrieval system for binding preparation, it is menu-driven and allows college and university libraries to retain a database of their periodical and monograph collections.

The software package was designed for use on most IBM-compatible personal computers. CompuSystem II simplifies record keeping, improves binding accuracy, contains help screens for each menu item, and is constructed with the ability to adapt to your present binder's slips.

For more information write CompuSystem II, 2101 N. Topeka Blvd., Topeka, KS 66608; or call 1-800-255-3502, in Kansas 1-913-233-4252. ■■

Minimal-Level LC-MARC Records Added to OCLC Online Union Catalog

Minimal-level cataloging records from the Library of Congress have been added to the OCLC Online Union Catalog (OLUC).

Between November 25 and December 4, OCLC added more than 135,000 LC-MARC minimal-level records of books published before 1987. New minimal-level records will be tape-loaded monthly beginning in 1988, and OCLC will process an estimated 40,000 such records annually.

With the addition of LC minimal-level records to OLUC, the OCLC database now provides online access to all LC-MARC records distributed by LC. The database contains more than seventeen million bibliographic records, approximately four million of which were provided by LC and thirteen million by input from OCLC member libraries.

Minimal-level cataloging records are for books whose use or research value is considered low by LC. The records are primarily for non-English-language materials; however, some English-language books and most microform collections in LC receive minimal-level cataloging. Minimal-level cataloging records also include brief romanized records for materials in Chinese, Japanese, Korean, Hebrew, Yiddish, Arabic, and Persian. ■■

SilverPlatter and Disclosure CD-ROM Databases

Information Access Company (IAC) has announced an agreement with two additional CD-ROM database distributors to offer their products on the new IAC InfoTrac Reference Center. SilverPlatter Information Services will provide the CD-ROM databases it masters and distributes for use on the new multimedia system. At the same time, Disclosure will provide subscribers with Compact Disclosure CD-ROM discs that can be accessed on the new system. Earlier, IAC announced that the Dialog OnDisc library of CD-ROM databases is available with the reference center. The initial Dialog product to become available for the multimedia network is the ERIC database.

The reference center enables libraries to construct reference networks with up to eight terminals providing access to as many as sixteen compact disc drives and eight videodisc players. In addition, the network allows for dial-out access to remote databases from all or any of the terminals on the system. IAC provides software and interface support for access to optical disc databases published and produced by other vendors. All of IAC's optical disc databases can be accessed simultaneously on the reference center. ■■

INNOVACQ in ARL Libraries

Boston University libraries have installed the INNOVACQ acquisitions and serials system, joining a growing list of ARL libraries using this system, including the University of Arizona; Arizona State University; University of California (Berkeley, Riverside, and San Diego); Dartmouth College; University of Kansas; University of Michigan; Michigan State University; University of New Mexico; Ohio State University; University of Oregon; and University of Virginia.

These large research libraries use INNOVACQ to perform fund accounting for materials budgets ranging from $2 million to more than $5 million. Their current serials subscriptions—in some cases over 30,000—are tracked by INNOVACQ's serials check-in and binding control programs. Some locations use up to thirty workstations and connect to offsite terminals via campus LANs.

All of the libraries employ online interfaces developed for use with INNOVACQ; OCLC or RLIN interfaces operate at every site. Boston University libraries and others electronically process thousands of books on approval with a unique online vendor interface. Many links with other local systems have been developed. The availability of such interfaces allows large, complex libraries to achieve a high level of productivity and to maximize the use of systems already in place.

INNOVACQ is available from Innovative Interfaces, Inc., of Berkeley, which has developed and marketed library automation products since 1978. Its local library systems and interfaces can be found in more than 150 libraries. ■■

Canadian Installations of LaserQuest

General Research Corporation recently announced the signing of four Canadian contracts for LaserQuest, a cataloging workstation that uses CD-ROM technology to distribute the GRC Resource Database of nearly five million MARC records. The Canadian libraries that will be using LaserQuest for retrospective conversion and current cataloging include those at Memorial University in Newfoundland and Laurentian University in Ontario as well as the Metropolitan Toronto Reference Library and the Government of the Northwest Territories, Public Library Services.

Memorial University selected LaserQuest for a retrospective conversion project of 500,000 records because of the high hit rates on large samples of the recon files. High rates are achieved with LaserQuest because it contains LC MARC records and more than two million records contributed by GRC customers in Canada and the United States for the past twenty years. It also includes foreign-language titles for books and serials as well as music and audiovisual materials.

Laurentian University selected LaserQuest for retrospective conversion and current cataloging. Diskettes of records downloaded from LaserQuest are loaded into its MultiLis system. A major conversion of sound recordings will be started this summer: university records will be matched against more than 132,500 music records. The university has a newly acquired map collection that also needs to be cataloged, and more than 36,500 map records are included on LaserQuest, which includes all MARC formats.

The Metropolitan Toronto Reference Library will be using LaserQuest for current cataloging. The Government of the Northwest Territories' Public Library Services provides current and retrospective conversion cataloging for twenty-two libraries. The Public Library Service also uses a LaserQuest link to its MultiLis system. ■■

Westchester Library System Purchases Utlas-Linked System T/Series 100

The Westchester Library System (WLS) has signed an agreement for the purchase of the Utlas T/Series 100 system. WLS is a cooperative public library system of thirty-eight public libraries serving nearly 867,000 residents, including Yonkers, New Rochelle, Mount Vernon, Greenburgh, Scarsdale, Mount Kisco, and North Salem, all located in Westchester County, New York. Thirty-seven of the thirty-eight public libraries have signed contracts to participate in the T/Series 100 system.

T/Series 100, a Tandem computer-based, distributed system, combines Utlas' integrated local system, the T/Series 50, with its bibliographic utility (Catalog Support Services/CATSS). The linked systems approach delivers cataloging with full authority control, plus complete access to the network cataloging database of 30,000,000 catalog records, to the local system. WLS' start-up purchase comprises 176 terminals and a six processor system (including Tandem's newest processor, the CLX), label processing, catalog services, implementation, and local system (circulation control and PAC) software. Installation begins in January and is expected to be completed in April 1989. The implementation will include creation of a local system database of close to 500,000 titles and five million items. ■■

CLSI Announces Three New Workstation Products for Cataloging and Offline Circulation Activities

CLSI, Inc., announced the addition of three new workstation products for use with their library automation system. The interface and cataloger's workstations create and maintain the library's bibliographic database; the offline circulation workstation records circulation transactions offline from the central computer. CLSI is the world's leading supplier of library automation systems, with over 300 systems installed and more than 1,500 libraries online.

The three workstations products operate on PC-AT compatible microprocessors, using the Intel 80286, 16-bit technology. The display monitor with 13-inch diagonal CRT produces high resolution characters in amber phosphor. The interface and cataloger's workstation software can share one hardware platform.

The interface workstation extracts MARC records from the major bibliographic utilities for downloading and storage in the LIBS 100 System. A "save" file program allows the library to store approximately 1,000 records offline. Each interface workstation supports up to four devices that can share the one interface.

The cataloger's workstation is a specially designed tool, allowing the library to add, modify, and delete MARC bibliographic and authority records. Cataloging procedures are streamlined through the use of prestored workforms, function keys, and default field values.

The offline circulation workstation records circulation activities while offline to the LIBS 100 System for use on bookmobiles, at branch libraries not yet online, and when the system is down. It can support up to four concurrent users. The workstation can also be used as an online terminal and can be configured with any CLSI scanning device for the scanning of barcodes. ■■

New York Regional Catalog Shows Nearly Three Million Holdings

When the members of the Southeastern New York Library Resources Council, Ramapo Catskill Library System, Mid-Hudson Library System, and the six school library systems put their holdings together, the result was a union COM catalog of 2,931,682 holdings.

The libraries had produced a union catalog earlier with another vendor but chose MARCIVE to generate the third edition. MARCIVE modified its unique index format to display information from the author, title, edition, publisher, and call number fields. In addition, the holdings were grouped by library system and central reference libraries marked with an asterisk.

Part of MARCIVE's COM service includes the production of a test catalog. This test catalog has already been approved, and final fiche production for over 250 libraries is beginning. ■■

LS/2 Sites Upgrading Their Systems

It has been a year since OCLC acquired

the ALIS I and ALIS II software from Data Phase and began providing hardware and software maintenance to libraries using these systems. OCLC renamed the software LS/2 and hired staff from Data Phase in order to effect a smooth transition.

Since that time, fourteen of the forty-six LS/2 sites have implemented system upgrades. Seven of the libraries have upgraded their central processing units (CPUs) to allow more users to access their systems, and another seven have expanded their disk storage to accommodate database growth. ■■

MELVYL Medline Funded for Year Three

The University of California has received a continuation grant of $160,236 from the National Library of Medicine to fund the third year of the MELVYL Medline project. This grant covers the final project period from April 1, 1988, through March 31, 1989, and brings the total amount awarded this project under a three-year Medical Library Resource Project Grant to $521,392.

The MELVYL Medline project will make the current three-year file of the National Library of Medicine's Medline database available throughout the University of California on the MELVYL online library catalog. Medline provides online access to article citations indexed from over 4,000 health sciences journals. It is a primary source of information in the life and health sciences for research, teaching, and patient care.

The MELVYL Medline project is being accomplished as a joint effort of the Division of Library Automation at the Office of the President and UC health sciences libraries. Coprincipal investigators are Mike Berger, assistant director, Division of Library Automation, and Mary M. Horres, biomedical librarian at UC–San Diego; they are being supported by a university task force of health sciences librarians and a users advisory council of health and life sciences faculty.

During the second year of the project, which is nearing completion, the MELVYL Medline database of 900,000 records was released for public access in UC's health and life sciences libraries.

In the third and final year, the emphasis will be on training and refinement of the system based on user evaluation. During this period, MELVYL Medline will also be made available in all UC libraries, extending access to users in all disciplines. ■■

CLSI Announces Installation of First Sequent System at Gainesville Public Library

CLSI has successfully transferred its LIBS 100 System software at the Gainesville Public Library in Florida to a new central processing unit manufactured by Sequent Computer Systems, Inc., Beaverton, Oregon. CLSI's new Sequent-based system uses a standard operating system, UNIX, which facilitates rapid software development through the use of standard programming tools. CLSI developed this new system to meet the requirements of large libraries and library consortia for automation systems that can operate many simultaneous applications and support a large number of terminals.

Using special migration software developed by CLSI's research and development department, the files and database of the Gainesville Public Library were transferred and converted from CLSI's proprietary operating system on Digital Equipment Corporation's PDP 11/44 processor to CLSI's new Sequent hardware platform.

CLSI and the Gainesville Public Library followed a detailed schedule for the system conversion. The Gainesville system was completely converted in a four-day transition period, during which circulation transactions were recorded offline on three CLSI back-up workstations and then subsequently loaded to the new Sequent system. ■■

OCLC Link Service Discontinued April 30

Effective April 30, OCLC will no longer offer its LINK Service, an electronic gateway for access to online databases from either an OCLC dedicated terminal or a dial-access terminal.

OCLC cited low overall usage and changing technology as the main reasons for discontinuing the service, which was introduced in November 1986. ■■

Recent Publications

Book Reviews

Kleper, Michael L. *The Illustrated Handbook of Desktop Publishing and Typesetting.* Blue Ridge Summit, Pa.: TAB, 1987. 770p. $49.95 (ISBN 0-8306-2700-6); paper, $29.95 (ISBN 0-8306-0700-5).

This massive book is as much about typesetting as about desktop publishing. It offers a great deal of good information in both areas, together with a certain amount of padding. While certainly not flawless, it is one of the best books dealing with desktop publishing.

The first chapter, "Type!," offers a fast-moving thirty-six-page blend of history and good advice on typography and page design—more good advice than many other books on desktop publishing offer in their entirety. "Word Processing as a Part of Typesetting" also includes much that is useful in its twenty-six pages, although some of the material appears outdated.

The next eleven chapters—550 pages—combine discussions of issues with profiles of particular programs and pieces of hardware. The chapters include word- and text-processing programs; tools for creating, generating, and reformating text; telecommunications methods; data manipulation and conversion; typesetting hardware tools (e.g., keyboard enhancements); professional typesetting software; specialized typesetting languages; desktop typesetting (98 pages); desktop publishing (200 pages!), and output devices.

Two short chapters (twenty-four pages combined) discuss typesetting decisions and business applications and appear wholly intended for traditional typesetting operations. The last chapter consists of lists of possible applications for desktop publishing and typesetting, and a number of brief user profiles. An eighty-four-page appendix lists various suppliers and offers concise profiles of programs; the book ends with a bibliography and what appears to be an appropriate index. There is no glossary.

The book is weakest in its program profiles and reviews. Kleper includes too many shots of Apple II menu screens; quite a few outdated programs for no-longer-available computers; too many very obscure programs; too many ugly dot-matrix "fancy" typefaces for the Apple II; and a lot of strange headline fonts for the Laser Writer and collections of "clip art" that can give publications all the personality of failing weekly shopping papers.

Kleper also, in more than one case, includes fairly lengthy profiles of two different versions of a program, old and new, which doesn't make any sense, unless he really expects readers to consider buying versions of programs that are no longer on the market. Many of the reviews are seriously outdated, and too many simply list menus and functions rather than discuss what the programs are all about. There are surprisingly few negative comments about programs.

This book will not replace a program-specific book on Ventura Publisher, Aldus PageMaker, or whatever desktop page production system you choose. It will not lead you through the specific steps of designing appropriate page layouts. But if you tune out the irrelevant and obsolete portions and ignore the Apple/Macintosh bias, you'll find lots of useful information and good advice threaded through specific program discussions and more general discussions.

This book also meets a criterion that many recent books on desktop publishing or page design don't: the book is, itself, well designed. While the pages are large, the text line is narrow enough for easy reading, and the typeface is a conservative, readable Roman, set fully justified with intelligent use of white space. The book doesn't "look designed," but reveals careful, thoughtful design making it easier to read and use.

Libraries should certainly have this book, although I can't suggest how the spine will hold up under the strain. It is one of the basic works on desktop publishing, and it also includes at least a book's worth of information on typesetting and related programs and techniques. I have not seen the hardcover version: a sturdy binding would be critical for this book.

Should you buy this book for your personal library if you're interested in desktop publishing? You could do a lot worse, and you could spend more money without getting as much useful information. It isn't perfect, but it's well worth the money.—*Walt Crawford, Editor* LITA Newsletter, *Principal Analyst for Special Projects, Research Libraries Group, Inc., Stanford, California.* ■■

Westerfield, Wiley. *Desktop Publishing/Teletypesetting: A Resource Guide to Electronic Publishing.* San Diego, Calif.: Westerfield Enterprises, 1987. 154p. paper, $16.95 (ISBN 0-942259-00-9).

Maybe you can't tell a book by its cover, but you can tell a lot about a book on desktop publishing from the first page or two of text. When the text is produced on a low-resolution dot-matrix printer with heavy letterspacing, it's hard to take it seriously.

In this case, you needn't bother. The best part of this book is the glossary, which takes up roughly a third of the book—and the glossary is marred by misspelling ("Helevetica") and a curious set of included words. (The author defines "Helevetica," Garamond, and Caslon, but not Bodini, Optima, Futura, or Palatino; he uses *masthead* within the definition for *flag*, but doesn't include an entry for *masthead* itself.)

Other than the glossary, there's surprisingly little here, and what there is leaves much to be desired. The book is a series of descriptions of products, ranging from one sentence to three pages. Longer descriptions read like advertising copy that wasn't checked for spelling, and the relevance to the title of the book is not always obvious. Prices are not included, and these are not reviews or coherent judgments. Each product includes numbers representing category codes, but there are no cross-indexes by category, only an alphabetical index by product name (which includes six listings under *The*).

This self-published book is apparently part of the series Tips for Self-Publishers. Here are three tips for free: don't waste good money on this book, check your spelling, and use something better than a dot-matrix printer, particularly if you're dealing with desktop publishing!—*Walt Crawford, Editor,* LITA Newsletter; *Principal Analyst for Special Projects, Research Libraries Group, Inc., Stanford, California.* ■■

Government Infostructures: A Guide to the Networks of Information Resources and Technologies at Federal, State, and Local Levels. Ed. by Karen B. Levitan. New York: Greenwood, 1987. 320p. $45 (ISBN 0-313-24864-8).

"Beneath any public policy and its government agency lies a supporting information infrastructure." This is the introductory statement in *Government Infostructures.* Just what does *information infrastructure,* or *infostructure,* as the Information Industry Association calls it, mean? Levitan, the editor, defines an *information infrastructure* as the underlying foundation of information resources and associated people, technologies, and facilities that supports decision making in organizations. She states that the book focuses on *government* infostructures because information is power, and no where else is this more evident than in the development of public policy. Levitan cites the questions that motivated the book as: What is the status of information resources used in the policy process? What types of information resources support the policy process at the federal, state, and local level? How are these resources shared across the levels of government and between the public and private sectors? Who has access to these resources and who does not? How are the resources established, maintained, and distributed?

The book is organized in three parts: the first chapters provide a description and discussion of the infostructures from the federal perspective; the second part focuses on state and local infostructures; and the third part is a series of case studies in selected policy areas.

Some years ago the U.S. Office of Technology Assessment conducted an evaluation of information technology practices across the federal government. One aspect of this assessment was a study of the impact of information technology on three civil liberty areas: access to agency decision making, collection of personal or company information, and the provision of public information. The first chapter, "Civil Liberties and Information Practices in Federal Agencies," summarizes the findings of that study. It includes a review of the laws, regulations and policies governing each of these civil liberty areas.

Monitoring and tracking legislation, legislative analysis, record-keeping, communication and oversight are some of the applications that Frantzich highlights in the chapter on "Use and Implications of Information Technologies in Congress." In the chapter on "Federal Data Systems: Policy and Practice," Griffith provides an overview of the development of federal statistical policies and a discussion of major issues currently confronting the statistical community; the issues include relevance, periodicity, quality, and the response burden. The last chapter of this section reviews current applications within the executive branch of a tool called personalized decision analysis (PDA). Brown, the author, defines *PDA* as an information technology that quantifies judgment on facts and values relevant to a decision and computes its applications for action. In all, the four chapters provide an excellent overview of the federal infostructures and the issues surrounding them.

The second part of the book provides a similar overview of infostructures on the state and local levels of government. Chapters focus on state legislative use of information technology, on information resources management practices in state government, on information resources management in Minnesota, and on American public opinion concerning privacy and civil liberties.

The chapter on public opinion is a synthesis of public opinion research; the conclusion drawn by the authors, Dutton and Meadow, is that privacy is a nonissue for most of the general public. However, the research also supports the findings that privacy is a latent issue and a class and civil libertarian issue. The findings are discussed in the context of five dimensions: (1) the concept of privacy; (2) the technologies linked to privacy; (3) the institutions associated with privacy; (4) the prospects for the future; and (5) the public policy responses.

The case studies of selected policy areas were some of the most interesting chapters because they clearly illustrated the relationships among federal, state, and local infostructures, the development of parallel infostructures in all political subdivisions, and the growth of infostructures around organizational goals, objectives, and programs. The National Data Base on Aging, described in the chapter on the aging network, is an example of a partnership between the public and private sectors. However, it also illustrates the problems that evolve when funding for an information network is shifted from the public to the private sector, when the government agency no longer has the ability to determine the type and quantity of information collected.

One of the case studies, nutrition policymaking in Hawaii, exemplifies how different infostructures develop even among organizations with similar interests when their clientele have different agendas. For example, the type of data considered important by those in charge of the food chain (farmers, processors, marketers) is not necessarily the same as that of the monitors of public health issues. It illustrates the tensions that exist among agencies in developing statistical programs and public policies. As the authors stated, "in retrospect, one wonders how there is a national nutrition policy at all."

Other case studies address decision analysis of postal automation, the Veterans Administration organization and information structure, hazardous waste management in Illinois and decision making in the Montgomery County (Maryland) school system. Each of these case studies provides a perspective on the organizational and other governing factors related to infostructures.

The essays in this book serve as a starting point for further discussions of information policies, technologies, and structures among librarians, politicians, administrators, students, and faculty. The essays are well written and include references for further reading. There is a good index, a list of acronyms, and biographical notes for all the contributors. For those who are interested in the dynamics of information and public policy, the book is highly recommended.—*Sandra K. Peterson, Documents Librarian, Yale University Library, New Haven, Connecticut.* ■■

Software Reviews

Inmagic. Inmagic, Inc., 2067 Massachusetts Ave., Cambridge, MA 02140-1338.

Inmagic is database management software that is especially suited for libraries, individuals, and organizations that must maintain lengthy texts on databases. It makes use of variable-length fields, and a given field may occur any number of times. Along with this feature, *Inmagic* comes with sophisticated retrieval and record formatting features that find a variety of applications in the library environment.

This software operates mostly in menu and prompt modes and, in the case of activities such as retrieval, in command mode. Most of the menus and commands can be invoked by just one keystroke. *Inmagic* provides extensive help whenever the user requests. In addition, a user can employ passwords that control access to the information in the databases.

When one is creating a new database, there is a logical progression of steps to follow: defining the data structure, opening a database file that is linked with the previously defined data structure, entering data, and retrieving information. One can also define one or more output report formats after the database structure is defined, but *Inmagic* can be used without going through this procedure.

DEFINING DATA STRUCTURE

Basic notions of defining the field structure of a database are not unique to *Inmagic.* When defining any database (for example, *dBASE III Plus*), the user must carefully decide how many and what fields are needed to describe the information at hand. There are, however, some unique aspects to *Inmagic*.

First, a given data structure can be used by any number of database files. The latter contain the actual information. Consequently, any change one makes in database structure, such as adding new fields, automatically affects every database file that is based on this structure.

Second, unlike in database softwares such as *dBaseIII Plus,* one does not have to specify the field length at all. This feature, obviously, suits people dealing with bibliographic information perfectly.

Fields are identified by two means: field labels and field names. For example, TI is a field label and TITLE is a field name. Field labels can be brief and cryptic and each field must have a label. Field names, in contrast, are optional. They can be longer and more informative. Field labels are embedded in the data and, consequently, once the data is entered into the database, labels cannot be changed. Field names, however, can be changed even after the database is created. If a field has both a label and a name, field name is displayed to the user. The user can use labels and names interchangeably. A database can contain up to seventy-five fields, and the total number of characters used for field labels cannot exceed 256 characters. The same is the case with the total number of characters used for field names.

Once a field is chosen, it is described to *Inmagic* by three characteristics: how it is indexed, how it is sorted, and how multiple entries of a field are to be treated when they are searched or sorted. In many cases, these descriptions are optional.

A field can be indexed in one of three ways: term indexing, keyword indexing, and both. Term indexing indexes the first sixty characters of a field and keyword indexing indexes every significant keyword within the field. In many cases, indexing decisions may not be crucial as the user can alter the indexing decision later on. Only the first fifty fields of a database can be indexed.

There are seven options in sorting a field—for example, letter-by-letter, word-by-word, leading articles ignored, numeric, and dates.

The emphasis code determines how a field with multiple entries will be handled by searching and sorting routines. One emphasis decision will make all the entries (for example, multiple authors in author field) fully indexed and searchable while another decision will make the secondary authors unindexed and unsearchable. The former decision is the default decision.

The user makes two other decisions regarding the fields—identifying a retrieval key and identifying the order in which the records are displayed after a search operation. The retrieval key uniquely identifies the record (e.g, accession number). The retrieval key can be a combination of two or more fields but the practice of combining two or more fields to derive a retrieval key should be avoided.

Assigning an order key, an optional operation, decides in which order records are displayed (e.g., by author and then by title).

Finally, the user can optionally assign passwords. A master password prevents unauthorized tampering with database structure. A password manager, the holder of the master password, can control the access to the database—for example, some users can only see the data, others can alter data in all or specified fields.

DATABASE FILE SPECIFICATION

In this case, the user assigns a database structure to a database file. A given database structure can be associated with a number of database files, but a database file can be associated with only one database structure.

DATA ENTRY

Data can be entered into the database in two ways: first, online, directly into the database—where the indexing takes place immediately after entering the data and saving the record—and second, in a batch mode, where one creates an ASCII file of one or more records outside the *Inmagic* environment and then downloads them into *Inmagic*. Batch processing may be convenient for organizing work flow. The same batch processing enables one to download records from online database systems such as Dialog.

SEARCHING THE DATABASE

Inmagic offers a host of sophisticated search and display options. Searches are conducted by using commands such as

GET AU EQ SMITH

where all the records satisfying the condition AUTHOR equals SMITH are retrieved. The format of a retrieval command is:

<command> <field>
<search relation>
<search term>

One can use any of nine search relations such as EQ, GT, GE, CW, and ST where these symbols represent EQUAL TO, GREATER THAN, GREATER THAN OR EQUAL TO, CONTAINING WORD, and STARTING WITH. In addition, Boolean operators AND, OR, and NOT can also be used as in the following case:

GET AU = PETERS AND PU = WILEY

The resulting records can be displayed on screen, sent to a printer or to a disk file. The records can be displayed in default order determined by the order key or can be decided ad hoc at the time of display. Records can be displayed in ascending or descending order by the sort key chosen. If one or more report formats are defined, they can be used in determining how the output will be displayed to the user.

One can carry out housekeeping operations such as displaying a directory of files or the structure of the databases, storing the searches into a work file, and viewing the keys in indexed fields.

REPORT FORMATS

Report formats control how the records in a given database are displayed or printed. For example, not every field in the database needs to be shown every time records are retrieved. Likewise, there may be occasions where complex outputs are required as in the case of a catalog card or a classified list of books and articles. *Inmagic* allows the user to determine highly complex output report formats, starting with the number of lines in a page, calculated fields, page layout in terms of report titles and column headings, and, finally, the most important feature, the record layout.

Report formats are described using an extensive report description language. For instance, a simplistic record layout is described as follows:

@LIST, LINE 1
IDNUMBER, BEGIN 'ID: '
@LIST, LINE 2
NAME, BEGIN 'AUTHOR: '
@LIST, LINE 3
TITLE, BEGIN 'TITLE: '

This description would simply print ID number, name, and title fields on three separate lines, identifying them by the literals ID:, AUTHOR:, and TITLE:.

For someone who is using *Inmagic* for the first time, the report-formatting feature is the most difficult one to master. Recognizing this difficulty, Inmagic, Inc. published guides such as *Biblio Guide*, which records and explains a number of database structures and one or more report formats based on them. For instance, *Biblio Guide* has report formats for catalog cards, authority lists, book and spine labels, and a number of others. These formats can be used with or without further modification by libraries and information centers. Or, using them as guides, one can design one's own specialized report formats.

Biblio Guide comes with a diskette that contains examples of six database structures and over ninety report formats that are based on those database structures. This means that time required to learn about creating report formats is substantially reduced.

TEACH OR HELP MODES

Another powerful feature of *Inmagic* is the availability of online help. There are explicit TEACH options where extensive tutorials are provided. For example, the TEACH command in database searching mode results in a menu of five options: Introduction to a Database, Searching Your Database, Using Boolean Commands, Controlling Searches, and Using the Results of a Search. Likewise, specific search commands are explained if one types in a command such as ? LIST—this command will explain the format and uses of the LIST command.

OTHER USEFUL FEATURES

Finally, there are other features that make *Inmagic* consistent and friendly. For example, *E* always stands for EXIT and many menus and search commands can be invoked by typing just one letter. Likewise, functions of keys such as F2 and F10 remain similar in all environments. These features, coupled with tutorial and help functions, make *Inmagic* easier to use.

HARDWARE

For greater flexibility, *Inmagic* should be used with a hard disk. It requires MS-DOS version 2.0 or higher and 384K RAM (both relatively common). It runs on a variety of mini- and microcomputers. This review is based on *Inmagic* version 7.0 running on an HP Vectra, an MS-DOS, IBM PC-compatible microcomputer.

PRICE AND OTHER RELATED INFORMATION

The first copy of *Inmagic* running on an MS-DOS microcomputer is priced at $975. Discounts are offered for buyers of multiple copies, and nonprofit organizations are eligible for additional discounts. *Biblio Guide* costs $145. Minicomputer versions of *Inmagic* are priced substantially higher. For an additional fee, *Inmagic* users can obtain telephone support and on-site training.

CONCLUSION

Inmagic is a powerful information management tool for small and special libraries and information centers. It can be used to develop a variety of applications, for example, for maintaining databases of patents, card catalogs, serials, audiovisual materials and other special collections.

Inmagic does not have a programming

language to learn and consequently does not allow the type of customization that *dBASE III Plus* allows, but one should remember that time and effort are required to write customizing programs. *Inmagic* cannot handle multiple database files at the same time, as *dBASE III Plus* can. Likewise, *Inmagic* differs from other popular softwares such as *PFS: Professional File* and Symantec's *Q&A*, in that these file managers are not terribly suitable for handling bibliographic data.

Inmagic has some disadvantages. The very attributes that give the system its sophisticated search features make it relatively difficult and unsuitable as a public access tool. Altering the database structure once the data are entered into the database file, can be done, but only with great difficulty and risk. Because of its complexity, learning *Inmagic* takes time, but the manual, *Biblio Guide*, and *Biblio Guide's* accompanying database structures, report formats, and online help features should be of great assistance. Although expensive and somewhat difficult to master, *Inmagic* is worth the investment of money and time in view of its flexibility and applicability to libraries.—*Rao Aluri, Tucson, Arizona.* ■■

Other Recent Receipts

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

Batt, Fred. *Online Searching for End Users: An Information Sourcebook.* Oryx Sourcebook Series in Computer and Information Science, no.1. Phoenix and New York: Oryx, 1988. 116p. $37.50 (ISBN 0-89774-394-6).

A Directory of Library and Information Retrieval Software for Microcomputers. 3d ed. Comp. by Hilary Dyer and Alison Gunson. Aldershot, England and Brookfield, Vt.: Gower, 1988. 75p. paper, $41.95 (ISBN 0-566-05586-4).

Dykema, Greg, and Dave Scholten. *Unix and Xenix Reference Guide.* Grand Rapids, Mich.: Abacus Software, 1988. 208p. paper, $12.95 (ISBN 1-55755-031-X).

Dykstra, Mary. *PRECIS: A Primer.* Rev. reprint. Metuchen, N.J. and London: Scarecrow, 1987. 262p. paper, $18.50 (ISBN 0-8108-2060-9).

Frontiers of Medical Information Sciences. Ed. by Robert Lawrence Kuhn. New York: Praeger, 1988. 228p. $39.95 (ISBN 0-275-92750-4).

Gervasi, Anne, and Betty Kay Seibt. *Handbook for Small, Rural, and Emerging Public Libraries.* Phoenix and New York: Oryx, 1988. 196p. paper, $27.50 (ISBN 0-89774-303-2).

Hartley, Jill, Amanda Noonan, and Stan Metcalfe. *New Electronic Information Services: An Overview of the UK Database Industry in an International Context.* Aldershot, England and Brookfield, Vt.: Gower, 1987. 147p. paper, $38.95 (ISBN 0-566-05489-2).

Kemper, Marlyn. *Networking: Choosing a LAN Path to Interconnection.* Metuchen, N.J., and London: Scarecrow, 1987. 279p. $37.50 (ISBN 0-8108-2031-5).

Kim, Choong Han, and Robert David Little. *Public Library Users and Uses: A Market Research Handbook.* Metuchen, N.J., and London: Scarecrow, 1987. 370p. $29.50 (ISBN 0-8108-2021-8).

Micro Software Evaluations, Library Edition. V.4, Ed. by Peggy Richard. Westport, Ct.: Meckler, 1988. 103p. paper, $95 (ISBN 0-88736-197-9).

Nationwide Networking: Proceedings of the Library of Congress Network Advisory Committee Meetings July and December 1986. Network Planning Paper, no.15. Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1986. 65p. paper, $7.50 (ISBN 0-8444-0573-6).

The New Information Professionals: Proceedings of the Singapore-Malaysia Congress of Librarian and Information Scientists: Singapore 4–6 September 1986. Ed. by Ajita Thuraisingham. Brookfield, Vt.: Gower, 1987. 357p. $77.95 (ISBN 0-566-05519-8).

Powell, Ronald R. *The Relationship of Library User Studies to Performance Measures: A Review of the Literature.* Occasional Papers Series, no.181. Urbana-Champaign, Ill.: Univ. of Illinois, 1988. 41p. paper, $3.00 (ISSN 0276-1769).

Software for Schools, 1987–88: A Comprehensive Directory of Educational Software Grades Pre-K Through 12. New York and London: Bowker, 1987. 1,085p. $49.95 (ISBN 0-8352-2369-8).

Webb, T. D. *The In-House Option: Professional Issues of Library Automation.* Haworth Library and Information Science Text Series, no.1. New York and London: Haworth, 1987. 166p. $34.95 (ISBN 0-86656-617-1).

Letters

To the Editor:

In A. B. Chitty's article published in the December 1987 issue there was appended a glossary, arguably useful for some readership but arguably unnecessary for *ITAL*'s readers. Be that as it may, Chitty's notion of what is a "clone" is certainly *not* the common usage. To characterize RLIN as "an OCLC clone" would by analogy make all OPACs clones of CLSI or LCS. It serves none of us to obscure significant differences, or is *ITAL* a clone of ________? —*C. James Schmidt, Vice-President and Director of RLIN.*

P.S. Cloning also presumes a chronological sequence, e.g., Ballots 19xx, OCLC >19xx. ■■

To the Editor:

Thanks for giving me a chance to respond to the letters published in the March 1988 issue.

I appreciate Douglas Livsey's comments: the ALOHA system came to my attention after the original analysis was complete and is quite an interesting product. Its response to ambiguous search terms can produce something like the browsing function I described. NOTIS functions in a similar way, and likewise the retrieval set is strictly limited in scope.

Stephen Salmon defends his Carlyle system well. As he knows, my own library has selected a Carlyle system for its OPAC. In late 1985, when I looked at various systems, Carlyle did not have an index browsing function; our own installation still does not, and I have not yet seen a Carlyle system with browsing implemented. In its absence, the heading indexes to which he refers can be created but are useful mainly to specify titles exactly word for word. Carlyle's keyword indexing of subfields is not generally different from other systems' approaches: the text is indexed if the profile specifies indexing for the field/subfield. This is not the function I described, in which the various types of subfields can be arrayed in a browsing order other than strictly alphanumeric, nor is it relevant in a system lacking a browse function like the Carlyle system circa 1985. Carlyle can, however, recognize and process indicators such as source in 6XX fields and can build separate indexes for separate topical controlled vocabulary systems.

I apologize for this error, which appeared in my original draft and was never corrected.—*A. B. Chitty, Library Systems Planner, Queens College Library, New York, New York.* ■■