Intellectual Access to Patron-Use Software

SHEILA S. INTNER

ABSTRACT

SINCE 1960, LIBRARIES EXPANDED from nonexistent collections of computer-based materials for patron use to having rapidly proliferating software and/or data file collections. Development patterns of patron-use software collections and the kinds of materials they may contain are identified. Bibliographic control issues for patron-use computer-based materials are explored, including the level of control required or desired, and differing forms of access. Responses to these issues by the library community are described, including development of standard tools for descriptive cataloging, indexing, and classification, and the application of standard bibliographic systems to two varieties of collections—*remote* and *local* access materials. Trends indicating future issues are outlined, and the author suggests that the best strategies are those that address control and access problems for the long term, although they may be more costly and difficult to implement in the short term.

THE RISE OF COLLECTIONS OF PATRON-USE SOFTWARE

In the three decades between 1960 and 1990, libraries and librarians have gone from having no collections of computer-based materials[1] for patron use and knowing very little about computers or the materials used with them to having rapidly proliferating software and/or data file collections,[2] or, where they do not exist yet, facing growing demands to establish them. The types of computer-based materials now available vary enormously in purpose, function,

Sheila S. Intner, Simmons College, Graduate School of Library and Information Science, 300 The Fenway, Boston, MA 02115-5898
LIBRARY TRENDS, Vol. 40, No. 1, Summer 1991, pp. 42-62

content, equipment required, and physical form. As a result, librarians must work hard just to keep abreast of a dynamic and increasingly complex marketplace of information products and services.

Library collections of computer-based materials have developed in different ways, depending on how and why they were initiated, which in turn was affected by the acquisition and use of computers elsewhere in the library's parent institution. Some of the earliest offerings at a few research libraries involved access to mainframe-based data files. These collections began to accumulate in sizable numbers early in the 1960s, and in 1974 the International Association for Social Science Information Services/Systems and Technology (IASSIST) was established by librarians handling them. Typically the data files in IASSIST libraries might have been generated locally within the institution or obtained from government or commercial sources, but they usually were not located within the library itself. Instead, the files were stored on the institution's mainframe computers, often located in a computer center physically and administratively separate from the library. To access the data on the mainframe, the library was given video display terminals without processing capabilities ("dumb" terminals) to use as input/output devices connected to the mainframe. For the users' convenience, the library retained any printed documentation that accompanied the data files, and it was the printed user guides, manuals, and other texts that comprised the library's part of the collection.

Beginning in the 1970s, libraries began to acquire access to mainframe-based bibliographic databases through membership in online bibliographic networks such as the Online Computer Library Center (OCLC) and the Research Libraries Information Network (RLIN), and by purchasing subscriptions to commercially distributed products purveyed both by nonprofit and profit-making organizations. Among the nonprofit, nonmembership bibliographic data systems to which libraries might subscribe were the National Library of Medicine's MEDLINE and the Library of Congress' MARC Distribution Service. Profit-making firms such as Lockheed and System Development Corporation (SDC) offered nonmembership bibliographic data systems that included large groups of individually produced online indexes and abstracting services previously available solely in printed book form—such as ERIC, AGRICOLA, and CHEM ABSTRACTS—which could be searched using one terminal and one set of commands. Databases such as OCLC, MEDLINE, and the DIALOG system were not considered library holdings, however, so librarians believed there was no need to exert bibliographic control over them, no reason to catalog and classify them, or to include bibliographic records for them in their catalogs.

At this time, libraries also began to purchase smaller computers—minicomputers—to perform library data processing such as inventory/circulation control and local catalog displays. Many of these minicomputers were part of a larger package of materials and services purchased from a vendor that included software as well as hardware, installation, training, ongoing maintenance and support, and research and development. Librarians did not consider the software they acquired in this way to be part of patron-use library holdings, either, and did not make any attempt to bring it under local bibliographic control.

Early in the 1980s, the development of microcomputers brought data processing into the realm of individual endeavor since, at least initially, microcomputers were intended as single-user machines. Unlike the larger minicomputers and mainframes designed to support group efforts with many disparate pieces of equipment—i.e., multiple inputs and/or outputs—microcomputers were self-contained units designed to process one task at a time for one input or output device (a typical library microcomputer station might have several input/output devices—e.g., keyboard, monitor, and printer—but the computer employs them one at a time). In addition to being smaller and single user oriented, microcomputers were inexpensive, hardy, relatively easy to learn machines that could be easily integrated into a library's existing environment. Microcomputers did not have to have specially controlled physical surroundings built for them, they could use ordinary electrical outlets, and their users communicated with them in English-language or quasi-English language style vocabularies. Librarians were quick to adopt microcomputing for administrative tasks such as word-processing, staff scheduling, personnel records, budget preparation, etc., and software began to be acquired to serve these purposes. As microcomputing became ubiquitous throughout society, in government, industry, and education, it is not surprising that libraries moved from staff-only software collections to patron-use collections as well.

The rapid spread of microcomputing elicited continuing research and development efforts to maximize the utility of the machines. Microcomputers continued to become smaller, faster, easier to use (termed friendlier), more powerful, and less costly. Most of all, methods of storing larger amounts of data—i.e., increasing the *microcomputer's memory—were sought. During the 1980s, new* products emerged that enabled microcomputers to store as much data as the minicomputers and, some say, the mainframes of earlier days, and to provide links between microcomputers and larger "host" computers located elsewhere. New linking products, such as local area networks (LANs), faster and more powerful modems, and

communications software, were adopted by libraries, although many of these were used solely for internal library processes. Among the new data storage products, hard disks that extend random access memory and pre-recorded laser optical disks called CD-ROMs (Compact Disk-Read Only Memory) were immediate hits with librarians. A hard disk attached to a microcomputer enables it to increase the amount of its local data storage by a factor of ten, twenty, or more beyond what their flexible magnetic disks could hold. This has altered the way software and data are being administered in libraries and will be discussed elsewhere in this article. CD-ROMs are a different kind of storage disk in which data are scanned by a beam of light instead of being read by a magnetic head. Their advantage is that a much larger amount of data fits on an optical disk than on the same size magnetic disk.

Data files on CD-ROM are a byproduct of the mainframe-based databases originally built for shared cataloging, such as OCLC, or for the production of periodical indexes and abstracts, such as those marketed in the DIALOG system—AGRICOLA, ERIC, etc. All or part of the online databases are recorded on a CD-ROM disk and marketed to the library for use in their microcomputers on-site. (The vendor usually will supply the special CD-ROM disk drive that must be attached to the microcomputer if it is needed.) Although the CD-ROM version of the database is static and does not reflect updates made to the database after it is recorded, it allows access to a more recent version of the database than printed book versions, and it enables the library to avoid the additional telecommunications costs incurred by direct online access to the database in the host computer. Libraries are a willing market for CD-ROM databases, purchasing subscriptions almost as fast as they appear on the market, hoping to give up-to-date, high-tech service to users at much lower costs than online access. For some unfathomable reason, subscriptions to CD-ROM databases are being perceived differently than access to their online counterparts, and librarians are trying to control them as they have always done for the printed book versions.

In the balance of this article, bibliographic control issues for patron-use computer-based materials are identified, responses to them by the library community are described and explained, and trends to watch as indicators of future developments are outlined, together *with this author's opinions on where attention might be directed* with positive results.

Issues in Providing Intellectual Access to Patron-Use Software

Establishing patron-use collections of any kind of material means

establishing some form of bibliographic control. Whenever a library wishes to control materials bibliographically, a host of familiar policy issues arise: What kind of control is required and desired—formal or informal? If a catalog of some sort is wanted, should full cataloging or brief listings be provided? What standards, if any, will be employed? How should the information be displayed for the library user? Who will provide it? These questions and others—such as how to process and under what conditions to store the materials, whether to allow open access to the shelves or other areas in which materials are kept, whether to classify the materials and how best to arrange them, etc.—have to be answered by librarians who decide to serve patrons with computer-based materials.

In the three decades between 1960 and 1990, two quite different types of patron-use collections of computer-based materials arose in libraries—mainframe-based data files and microcomputer-based software and data files. (Minicomputer-based software generally was used solely for the library's internal data processing and was rarely documented in the same manner as patron-use materials.) The mainframe-based data files, which began to be collected early in the period, were stored on computers located far from the point of use. The files were supplied to end-users in the library via terminals with textual documentation kept in the library. Thus, the "materials" themselves were invisible, both to the librarians and the users. In contrast, the microcomputer-based software and data files, which began to be collected in the last third of the period, usually were stored on site in the library. They received treatment similar to that of other nonbook materials such as sound and videorecordings, and might be used in the library itself with library-owned microcomputers, or, they could be borrowed in the same manner as books and other library materials and used elsewhere by the end-user. The two types of collections, which came to be called "remote" and "local" access materials, elicited different perceptions about how to treat them bibliographically, as described earlier.

In some libraries, microcomputer software and data file collections began as small numbers of titles intended for staff, but the size and staff-only focus changed quickly as the potential for patron service was recognized and addressed. When collections and user groups were both small, collection control could be informal, in the form of simple lists. The availability of thousands of titles within a short time after the introduction of microcomputing, however, and libraries' desire to acquire more and more titles plus their shift toward patron use of materials made it difficult for them to continue controlling rapidly growing collections so casually. In some places, centralized microcomputer laboratories were developed

for patron use where computer-based materials were stored and administered. Depending on the orientation of laboratory administrators and the expertise of their staff's, control might continue to be informal or locally-devised. (Reports from all types of libraries appearing in the literature attest to librarians' inventiveness in devising local treatments for computer-based materials [Baker, 1985; Dumlao & Cook, 1983; Mead-Donaldson, 1984].) Elsewhere, microcomputer hardware and software (including data files) were distributed throughout the institution. In either type of setting, however, the need for formal systematic controls arose in order to let users know what materials were available (i.e., to provide access) as well as to keep track of the holdings (i.e., to maintain bibliographic and inventory controls).

Ongoing rapid developments of hardware, software, and data file storage technologies make it difficult to identify short-term, simple, quick, and easy solutions to problems of access and control. The best strategies seem to be those that address access and control issues for the long term, and that acknowledge a need for flexibility and the development of staff with expert knowledge, even though such solutions tend to be more costly and difficult to implement in the short-term. In the next section, standard methods for bibliographic control and patron access are described.

Organizing Patron-Use Software: Rules and Tools

Organization of library materials rests on three components for which the library community has developed standards: description and access, indexing, and classification. Also, in view of the computerization of current library operations, the ability to transform bibliographical data into machine-readable form is assumed and standards for it should be added to those for the three components of bibliographic control and access. Standards accepted in the United States library community for these elements of organization are:

- the *Anglo-American Cataloguing Rules* (Gorman & Winkler, 1988; 1978) for description and access;
- *Library of Congress Subject Headings* (Library of Congress, 1990) for subject descriptors that comprise the indexing vocabulary;
- either the Dewey Decimal (Comaromi et al., 1989) or Library of Congress classifications (Library of Congress, 1917);
- and, the *MARC Format for Bibliographic Data, Computer Files* for machine-readable coding of data.

Each standard and its development are discussed later in greater detail, but it should be understood at the outset that there are no intrinsic obstacles to applying these standards to library software and data files. Furthermore, in this author's opinion, the advantages that obtain

from following uniform, standard, mainstream practices for books and other types of materials acquired for patron use apply equally to computer-based materials.

Cataloging of Computer-Based Materials Using AACR2R (the 1988 Revision to AACR2)

The first chapter of AACR2R, the standard code for describing materials and formulating headings based on descriptive elements, includes rules for all materials currently collected in libraries, including computer-based materials. AACR2R's rules are based on the international family of standards developed by the International Federation of Library Associations and Institutions (IFLA), known as International Standard Bibliographic Description (ISBD), which mandates the sources from which data should be taken, the elements to be included in the record, the order of elements, and the punctuation used to identify them.[3] In addition to the first general chapter, AACR2R's chapter 9, titled "Computer Files," contains special rules that apply solely to computer-based materials, defined as "files that are encoded for manipulation by computer . . . data and programs . . . stored on, or contained in, carriers available for direct access or by remote access" (Gorman & Winkler, 1988, p. 221). The scope statement goes on to instruct catalogers to use chapter 10, "Three-dimensional Artefacts and Realia," for cataloging electronic devices such as calculators or software residing in a computer's permanent memory (i.e., ROM), which is considered part of the piece of equipment (Gorman & Winkler, 1988, p. 221).

Following this admonition, the rules themselves are quite similar to the rules for other types of materials found in other chapters, and only the unique features are addressed here, arranged by the element or area of description to which they relate (parenthetic numbers refer to related rules in AACR2R).

Data Sources (9.0): Data sources are adapted to the availability or lack of availability of computers to run the item being cataloged. Title screens are the preferred chief source of information, but, acknowledging that they are not a viable data source for catalogers without appropriate hardware, information from permanently affixed labels on carriers (i.e., disks, tapes, cartridges, or other storage media), accompanying documents, or containers (i.e., boxes or other disposable packaging) may be substituted in that order of preference. The source of the title must always be noted to aid the user of the catalog record in identifying items that may have different titles in the various locations.

Title and Statement of Responsibility (9.1): Catalogers are cautioned against using file names or data set names as titles, unless

they are the only names available. Another instruction directs that sponsors be listed in the notes rather than in the statements of responsibility.

Edition (9.2): Terms indicating edition are augmented to include "version," "release," and "update," popularly used for computer-based materials in place of "edition."

Material Specific Details (9.3): Information about the character of the files is contained here, including whether they are data and/or programs, the number of individual files, and their length or composition. These data are particularly important to searchers trying to identify remotely accessed files.

Publication, Distribution (9.4): Computer file producers are equated with publishers.

Physical Description (9.5): Most of the special rules pertain to describing the physical composition of the item. Terms for carriers, attributes such as sound and color, and specific instructions about how to record dimensions for different types of carriers are related to various storage technologies.

Series Statements (9.6): No special rules appear for this area.

Notes (9.7): In addition to the special note for the data source used for the title, unique notes include "system requirements" (describing the hardware and other requirements for using the item), "file characteristics" (adding information not contained earlier in the record), and "other formats" (in which the issue of the same file for use with another type of computer may be noted). Also, under the instructions and examples for data pertaining solely to the copy of the item being cataloged is the direction to record a data set name (and presumably, although it is not specifically stated, a file name), if desired.

Standard Number and Terms of Availability (9.8): No special rules appear for this area.

These rules, which seem to work well when the catalogers applying them have sufficient familiarity with computers and computer-based materials to understand them, evolved from an earlier version of chapter 9 published in the original AACR2 in 1978. The rules in the earlier chapter were based on an assumption that no physical item in hand was possible since the materials—data files—would actually reside on a mainframe at some distance from the point of use (or the point of cataloging). They had no provision for physical description, and the number, size, and other characteristics of the invisible files were substituted for a description of physical objects. Information was expected to be taken from documentation rather than from the files themselves. The materials were called "machine-readable data files," failing to acknowledge that libraries might collect

programs, program packages, or the other items more generally termed *software.*

"Machine-readable data files (MRDF)," the official designation for these kinds of materials, was challenged on four counts: (1) for being too lengthy, particularly since cards still were the dominant catalog display medium; (2) for being too narrow in scope, since it did not appear to include software; (3) for being slightly inaccurate, since microforms, motion pictures, videorecordings, etc., all could be deemed "machine-readable" as well as computer files; and (4) for failing to include "computer" as one of the words in the phrase, which caused confusion for persons unfamiliar with the medium.

Lobbying efforts to change the name and the focus of the 1978 chapter from remotely stored mainframe files to locally available microcomputer materials began with the advent of microcomputers in libraries along with collections of microcomputer software. Although data file librarians protested that "data file" could be construed to include software, and that files were files whether they were stored on a mainframe or on a disk that one put into a microcomputer, it became clear as time passed that a groundswell of dissatisfaction with the then-current rules was gaining momentum, and that it could not be ignored.

National level groups in each of the countries responsible for AACR2 began working on alternative rules. In 1984, after eighteen months of work on the part of a dedicated task force co-chaired by Arnold Wajenberg of the University of Illinois and Ben Tucker of the Library of Congress, the American Library Association's Committee on Cataloging: Description and Access (CC:DA)[4] approved and published an interim interpretation of chapter 9 titled *Guidelines for Using AACR 2 Chapter 9 for Cataloging Microcomputer Software* (ALA, 1984). This brief document, which had authority solely within the U.S. library community, explained such arcane exercises as how to count the files in a program package contained on a 5.25 inch floppy diskette and what to do if the number could not be determined, added a means of identifying the number and type of disks or other physical objects deemed "carriers" of the computer files being cataloged, and created a "systems requirement" note to describe the hardware needed to use the item. The guidelines did not alter the general material designation or define the material specific details area. It included a relatively large glossary of computer terms that library catalogers found very helpful, since many of them were not conversant with the jargon.

Similar efforts abroad resulted in reports from interested groups in Great Britain, Canada, and Australia, and, eventually, an official proposal from the British to the Joint Steering Committee for

Revision of AACR (the international body with sole authority to make rule revisions to AACR) for changes to chapter 9. The British proposal stimulated CC:DA to appoint a new task force to consider methodology for reviewing the chapter preliminary to formulating a U.S. proposal or responding otherwise to the need for revised rules. During the same period, IFLA had appointed a working group charged with creating an ISBD for computer files which would be completed and published in 1989, a year after the AACR2 revision (ISBD [CF], 1989). As these various groups worked on reports and proposals, gathering information, sifting it, and weighing alternatives, AACR2 editor Michael Gorman drafted a new chapter 9 and worked feverishly with members of the Joint Steering Committee to obtain unanimous approval for it before the publisher's deadline of December 1986. At that time, the publishers were going to press with a revised edition of AACR2 that would incorporate the many large and small changes to all parts of the text authorized since 1978.

Final approval of a new text of chapter 9 was obtained from the Joint Steering Committee in late 1986, and, acknowledging the need, it was published separately (Gorman, 1987). The text appeared, with minor amendments, in 1988 in the new issue of AACR2 and was accepted as the current standard at this writing. The most dramatic changes were the following:

- the chapter name and general material designation was changed from "machine-readable data file" to "computer file"—a compromise;
- data sources were made consistent with other chapters and with the principles of preferring sources closest to the item itself;
- information about file characteristics were removed from physical description and relocated to the area for material specific details, newly defined for computer files;
- physical objects in hand—i.e., the disks, cartridges, etc.—called "carriers" were described in the physical description area in the same manner as for all other types of material; and
- notes, such as systems requirements, file characteristics, etc., were augmented and interpreted appropriately for computer-based materials.

One might believe that nine or ten years is a long time to accomplish the rule changes in AACR2 that catalogers needed to describe microcomputer software easily and adequately according to an authoritative standard, but for any endeavor involving several countries and diverse constituencies, the time frame probably is not unusual. The 1988 standard code—i.e., AACR2R—makes description and descriptive access for computer-based materials consistent with

all of the other types of materials covered by the rules, and affords librarians the valuable opportunity of integrating these records with records for books, maps, sound recordings, films, videos, and everything else cataloged in the operational mainstream.

Indexing Computer-Based Materials

To follow the standard procedures for indexing (called "subject cataloging" by library catalogers), terms used as subject descriptors for computer-based materials must come from whatever authorized list of terms is used for other materials, namely *Library of Congress Subject Headings* (LCSH), *Sears List of Subject Headings* (Sears), *Medical Subject Headings* (MeSH), or another published standard indexing vocabulary used by individual libraries. There is greater tolerance for variation among libraries in adopting a standard for indexing than is acceptable for descriptive cataloging, because, ideally, the vocabulary chosen should match users' capabilities—i.e., the knowledge levels and subject expertise of individuals using the catalog—as well as accommodating variables of collection size and degree of subject specificity. Since collection size, scope, depth, and user profiles vary from library to library, one standard vocabulary is unlikely to satisfy them all equally well.

The most widely used of the three lists mentioned earlier, at least within the United States and Canada, is the *Library of Congress Subject Headings* (LCSH). A majority of academic and public libraries use LCSH, as do large numbers of school and special libraries. LCSH's current popularity may well be attributable to its use on printed catalog cards distributed by the Library of Congress since the early 1900s and the availability of the published list dating back to 1909. Even if they were not perfectly matched to a library's needs, LCSH descriptors were there for the taking, saving individual libraries the time, effort, and cost of purchasing and using another tool solely for subject descriptors or devising and documenting descriptors of their own. In view of its wide use among the several sectors of the library community, only LCSH will be described here. Much of the discussion is applicable also to Sears (published by the H. W. Wilson Co.) and MeSH (published by the National Library of Medicine).

Nine complete editions of LCSH appeared by 1980. In the decade from 1980 to 1990, however, this standard tool containing more than a quarter of a million descriptors was transformed from irregularly issued editions of printed books—the familiar large red volumes—or more frequently issued microform versions, to a fully computerized online file available in any of several computer-based media including magnetic tape, CD-ROM disks, and direct online service for any library linked to the Library of Congress as well as in microforms

or printed books, now issued annually. The task of digitizing the list was a formidable effort, requiring first the development of a MARC format for subject authorities, a template that could accommodate authorized heading forms, unused forms, several kinds of references, and documentation of sources, and subsequently a massive project inputting the records for each of the hundreds of thousands of existing descriptors. Once completed in 1989, however, the online file (called LCSH-mr for "machine-readable") became simple to change in order to add, delete, or alter descriptors.

Since the early 1980s, totally new descriptors for books about computers and all the sub-topics in computing have been established in large numbers that show no sign of diminishing. If anything, problems might occur now because too many potentially overlapping terms are authorized in LCSH—e.g., "Computers," "Electronic digital computers," "Minicomputers," and "Microcomputers." At one time, Library of Congress subject catalogers resisted accepting new terminology rapidly or changing established terms to update terminology—e.g., "Electronic data processors" to "Computers"—because of the cost and staff time required to alter existing records, but with the additional flexibility afforded by the online status of both bibliographic and subject authority files this barrier is rapidly disappearing.

The pilot project for the Machine-Readable Collections Reading Room (MRCRR) at the Library of Congress has been made permanent. Since the autumn of 1989, all software producers are required to deposit a copy of their items with the Library of Congress; LC is no longer willing to catalog from the documentation alone. The free floating subdivisions of "Software" and "Juvenile software" are now being used for all packages cataloged by LC, and 10-15 headings for the software itself are now being reviewed (D. Beaubien to Ann Fox, cataloger, Special Materials Cataloging Division of the Library of Congress, personal communication, January 23, 1991).

Using a list of descriptors that contains relevant terms is the first and most important concern, but it is not the sole concern for librarians who wish to provide effective subject access. The second concern is the way the descriptors are applied—i.e., the policies governing their use. The Library of Congress' Subject Cataloging Division makes such policies for its own operations, and, at catalogers' requests, began publishing these policies for general use in other libraries (Library of Congress, 1990). Unfortunately, computer files were not routinely collected and cataloged at the Library of Congress until the summer of 1988, so few policies for their subject access had been established before that date. The library approved a pilot project to provide Cataloging-In-Publication for 1,000 computer files

early in 1987, but implementation was delayed because completion of higher priority projects has taken longer than anticipated. Thus, in considering LCSH descriptors for use with computer files, one must understand that little help is forthcoming from the usual sources. Also, LCSH descriptors were devised solely with books in mind. While many descriptors are valid for topical information in any physical format, all of them are not equally appropriate for books *about* computer software and the software *itself,* which are two quite different things.

Policies for effective access to nonbook materials have long been a concern of librarians in the field, and two committees of the American Library Association's Association for Library Collections and Technical Services have addressed them—the division's Audiovisual Committee and the Subject Analysis Committee of its Cataloging and Classification Section. Joining forces to help the Library of Congress develop plans for its microcomputer Cataloging-In-Publication project, the two committees have worked both cooperatively and separately on assisting librarians with problems of subject access. As early as 1984, the Subject Analysis Committee appointed an ad hoc subcommittee to propose guidelines for subject access for microcomputer software and held hearings to solicit ideas, opinions, and responses from librarians. In 1986, the recommendations of the subcommittee were published, furnishing four fundamental principles and one caveat to be followed in making local policy decisions for the subject cataloging of software (*Guidelines on Subject Access...,* 1986).

- treat microcomputer software in the same manner as all other materials;
- assign subject descriptors and classification numbers using the same standard tools as for other materials;
- use the same criteria to determine subject content and represent it in descriptors and classifications, generally classing first by topic, then by form;
- if a form subdivision is desired, the term *software* is suggested;
- do not make main headings for the form of the software or for the make/models of the hardware, operating systems, etc., although they could be subdivisions (pp. 5-6).

For the most part, these principles are based on common sense and the desire to use descriptors to reveal the *subject content* of materials. One can visualize easily how useless the suggested subdivision "software" or any other form heading would be to searchers if it was the primary descriptor for numerous items covering topics from arithmetic to zoology. Following the principles is not difficult if catalogers can determine the subject content of the materials

they are indexing. The lack of requisite hardware to sample the contents of a piece of software or the failure to understand descriptions of its contents and intended uses because of unfamiliarity with the terminology both may confound effective indexing every bit as much as the failure of LCSH to contain a desired descriptor.

Classifying Computer-Based Materials Using Dewey and LC

Issues discussed in connection with LCSH apply also to the use of standard classifications such as the Dewey Decimal Classification (DDC) and Library of Congress Classification (LCC) for computer-based materials. Both classifications were devised with books in mind, and their categories and terminology are not always appropriate for nonbook materials. Often, desired categories are missing because particular topics appear in computer-based manifestations before they surface in books, and, without a basis in printed books (i.e., "literary warrant"), the topics will not be established by the classification. The Library of Congress, which assigns "official" DDC and LCC numbers to books, does not classify computer-based titles and thus does not supply even a nucleus of examples for other catalogers to emulate, nor does it establish policies for their classification (in OCLC, approximately 32,000 records for computer files have been entered into the Online Union Catalog, only four of which are attributable to the Library of Congress). Nevertheless, it is incumbent upon catalogers to arrange computer files in some meaningful order, especially if local policies mandate open stack patron-use collections that lend themselves to being browsed.

Both DDC and LCC are enumerative classifications and are based on the principle of classification by discipline. The enumerative character of the classifications means that, to be assigned, classes must be available in their schedules. Missing classes cannot be constructed by the classifier when they are needed. The disciplinary based character of the classifications means that, in both schemes, materials about different aspects of computing will be classed far from one another rather than being collocated in one place on the shelves. In DDC's 20th edition (1989), most computer-related topics occur at 004-006, but research and information systems are at 621+ with electronic engineering, and 519+ with mathematical probability theory. The 004-006 schedule is a complete and greatly expanded revision of numbers at 001.64+ in the previous edition, which were filled to overflowing with the outpouring of computer-related publications of recent years. The few classes available in the 19th edition could not organize and arrange the rapidly developing subject area. Listings in the Relative Index for terms beginning with the

word "computer" also include numbers in the Social Sciences (3xx), Business (65x), Printing (68x), Art (700), Games (79x), and Library Science (025+) as well as numbers in the auxiliary tables in which some aspect of computing is to be added to class numbers from the schedules.

LCC also separates materials about computers by disciplinary focus, placing most topics at QA76+ as a subset of mathematics and at TK (electronic engineering), although other topics may be found elsewhere, too, depending on the aspect of computing being represented. Unlike DDC, LCC does not have a combined index to which one can turn for a collocation of terms beginning with the word "computer." It is more difficult to see an array of classes for computer-related topics in one place. But LCC's much greater specificity, reflected in a much larger number of classes, accommodates close classification more easily than DDC. LCC also expanded its principal computer-related sections in QA and TK considerably to accommodate new topics and topics requiring additional subdivision.

A major difference between DDC and LCC is the way they subarrange materials within a more general class. DDC is hierarchical and tries to place topics in meaningful relationships to one another. LCC is not hierarchical and usually leans toward alphabetic or geographic subarrangements (and, sometimes, both together). While LCC's arrangements are systematic and organize large collections with relative ease, they are not satisfying to browse, since materials are interfiled without regard to their subject relationships.

An important adjunct to classification numbers in completing a shelf address for each item in the patron-use computer-file collection is the assignment of book or shelf marks. Shelf marks (it seems counterproductive to call them "book" marks when the focus of the discussion is not books) may include Cutter numbers; dates; collection marks such as "Reference," "Branch," or "Juvenile"; and volume numbers and/or copy numbers, depending on local library policies. Dates, collection marks, volume, and copy numbers are as easy to assign appropriately to computer files as to any other type of material. But Cutter numbers are another thing, and greatly expanded lists of Cutter numbers have been devised and published that offer valuable assistance to catalogers dealing with large computer file collections (Leysen, 1986a, 1986b).

There are compelling reasons to utilize the same classification and shelving systems for patron-use computer materials as are used for other patron-use materials. Adoption of the same classification for all library materials regardless of their physical form enables both patrons and staff to transfer what they know about the subject classification and arrangement of one type of material to all other

types. This makes computer-based titles easier to classify for members of the cataloging department staff, easier to shelve for members of the collection maintenance staff, easier to retrieve for members of the reference staff, and easier to browse for members of the public. There do not seem to be any compelling reasons to do otherwise, for even if the collection is completely closed, the classified arrangement can be made available to searchers through an index or shelflist. In the event the collection is closed to patron browsing, assistance in selecting an item rests entirely on the catalog record, including the subject descriptors and classification that might appear there. Given the useful nature of classification for browsing and serendipitous discovery, it would seem a shame to eliminate these potentials for service.

Coding and Tagging According to the MARC Format for Computer Files

The final element in standard bibliographic access and control is inclusion of the bibliographic records in a computerized database of bibliographic information in a standard format. Lack of entry into such a database means exclusion from the mainstream of library materials and services since computerized bibliographic networks have become the most important sources of information for collection development, cataloging, and use, and local library systems usually depend on the availability of data in this form. The MARC (*MA*chine-*R*eadable *C*ataloging—i.e., USMARC) format developed by the Library of Congress has become the U.S. national standard communications format for computer-based bibliographic data, not only by default, since no other standard has been developed, but also by virtue of its publication as a standard of the American National Standards Institute (American National Standards Institute, 1977; Library of Congress, 1980).

Separate formats have developed over the years for monographic books, serials, films and videos, musical scores, sound recordings, maps, and other types of materials. A MARC format for computer files (and its predecessor, machine-readable data files) was, indeed, developed by the groups responsible for the standard—i.e., the Library of Congress' MARC Development Office and the American Library Association's interdivisional Committee on Representation in Machine-Readable Form of Bibliographic Information (MARBI).[5] As mentioned earlier, the format has not yet been made available for use by catalogers at the Library of Congress at this writing, but it has been adopted and implemented by the major bibliographic networks and is being used by the thousands of libraries cataloging in those systems.

Details of the computer files format are similar to those of other

formats with some exceptions. The coded description of the item being cataloged, known as the "fixed fields" in OCLC and RLIN and the 008 field in USMARC, has special fields to represent the type of files being cataloged and the type of machine they require (i.e., a computer or "other" type of machine), features unique to computer files. Two special variable fields, 538 and 753, accommodate the systems requirement note and an added entry for the make and model of the computer, respectively. Fields for title variants accommodate computer files' penchant for having acronymic "official" titles that are spelled out subsequently or spelled out "official" titles that are acronymized elsewhere on the items. Fields for some of the information associated with serial publications are defined in the computer files format so that serially published files can be represented without having to substitute use of the serials format, which in turn would not be able to accommodate some of the data unique to computer files.

In recent years, dissatisfaction with the proliferation of separate formats having differing field definitions and the ensuing inconsistencies among formats resulted in calls for integration of the formats into one consistent structure (Attig, 1983; 1989). In 1987, a format integration proposal was put before MARBI and agreement on various issues resolved in the years that followed. Knowledgeable experts suggest that implementation of the final format integration proposal will begin in the field before 1995.[6] Format integration may have fewer impacts on the computer files format than, for example, on the audiovisual/visual materials format used for films, videos, etc., because fields for representing certain types of data such as multiple name versions, seriality, etc., have already been defined in the existing computer files format. Since the computer files format was a recent addition to the family of MARC formats, developers were conscious of the problems created by failure to include these fields in the formats for other types of media.

Current Status

To sum up, standard tools supporting standard policies and practices for cataloging, indexing, classifying, and computer coding information for access to and control of computer files are firmly in place and as fully developed as they have ever been for any nonbook media materials. That they should be employed in place of nonstandard alternatives for providing access to patron use data file and software collections is highly recommended without reservation by this author. Nonstandard alternatives, no matter how attractive they may appear to be, are not part of the mainstream of library systems and services and stand to fail to remain adequate over the long term for two reasons:

1. nonstandard alternatives eventually add more work to information storage and retrieval processes than standard treatments by requiring special tools, training, and systems; and
2. nonstandard alternatives eliminate the potential for economy, effectiveness, and efficiency by being incompatible with mainstream systems and services.

For the Future

Two factors indicate that the delicate equilibrium suggested by the foregoing conclusion is unlikely to persist for very long. The first and perhaps the most important factor is that technological developments in computing are being pursued vigorously and are intensely competitive, resulting in a field that is not just dynamic but highly volatile. Also, it is a field over which librarians exert very little, if any, influence. The second factor is that library responses to progress in the field have, to date, tended to be extremely slow, cautious, and limited, focusing on making as few changes as possible. This kind of scenario tends to lurch from crisis to crisis without much hope for developing a flexible, responsive, knowledgeable, and reliable problem-solving structure. Some of the potential crises that loom ahead include the following:

- the use of microcomputers with hard disks means librarians can load microcomputer software onto hard disks and distribute them via local or wider area networks, and implications for bibliographic access appear to be very similar to the original mainframe-based data files in which no item in hand was available for cataloging, indexing, and classification;
- interactive multimedia technologies employ computer software, video, sound, and textual images with user responses in new combinations that have not yet been addressed by access tools;
- new emphases on resource-sharing and cooperative collection development projects would indicate the greater importance of access to materials held outside the home library, but shared by it; and finally
- new products and services are tilting heavily in favor of full-text, cataloged, and indexed collections of titles in online or CD-ROM databases as opposed to individual items marketed separately and will require entirely new kinds of policies and treatments.

Clearly, librarians need to undertake a thorough exploration of options available for expanding access to local materials as well as for establishing links with collections outside the library and the parent institution, and they need to do so quickly. Simpler mechanisms for altering cataloging rules, lists of subject descriptors, and classification schedules must be sought to help librarians meet

the demands of a rapidly evolving field. Methods of incorporating user knowledge into the process, suggested by Bates (1989) and others need to be considered and addressed (Koenig, 1990). Strategies for utilizing the cataloging that might accompany purchased products and integrating it with other library cataloging need to be anticipated. At the same time, librarians must prepare themselves with more knowledge, better training, and more flexible managerial skills, including critical evaluation, decision-making, risk-taking, problem-solving, and creative thinking, and build staffs with more knowledge, better training, and a desire to contribute to patron service in order to meet the tests that lie ahead.

ENDNOTES

[1] For the purposes of this paper, the term *computer-based material* includes all informational materials requiring a computer to use, read, view, or hear.

[2] For the purposes of this paper, the term *software* will include individual programs and groups of programs known as program packages that enable people to use computers to perform various processing manipulations. The term *data file* will include any type of textual or numeric data (with the exception of programs or groups of programs) requiring the use of a computer to read, view, or hear—e.g., the word-processing package, Microsoft Word used to write this article is a software item, but the actual text is stored separately on a computer disk and is considered a *data file.*

[3] The ISBD structure calls for the following elements in this order: Title and statement of responsibility; edition; material-specific details; publication, distribution information; physical description; series statement; notes; and standard numbers and terms of availability. The newest of the ISBDs, developed originally for monographic books (ISBD[M]) and subsequently for serials (ISBD[S]), printed music (ISBD[PM]), and other material forms is (ISBD[CF]) (for computer files).

[4] The position of the Committee on Cataloging: Description and Access in the organizational hierarchy of the American Library Association is not as simple as it would appear from this statement. In the early 1980s the committee (abbreviated CC:DA) was part of the Cataloging and Classification Section of the Resources and Technical Services Division of the American Library Association (i.e., CC:DA/CCS/RTSD/ALA). In 1990, the division changed its name to the Association for Library Collections and Technical Services (ALCTS). Thus the committee, which remains at the fourth level of hierarchy, could now be fully abbreviated CC:DA/CCS/ALCTS/ALA.

[5] MARBI consists of three representatives each from the Association for Library Collections and Technical Services, Library and Information Technology Association, and References and Adult Services Division. In addition the committee has ex officio representatives from the Library of Congress and the National Library of Canada as well as receiving liaison representatives from OCLC, the Research Libraries Group (RLIN), Utlas International, the Western Library Network, CLASS, NOTIS, UCLA (Orion), the U.S. Government Printing Office, the National Agricultural Library, the National Library of Medicine, and other interested groups.

[6] This statement is based on reports of OCLC's Glenn Patton and RLIN's Ed Glazier to the membership of OnLine Audiovisual Catalogers given at their business meeting on 18 October 1990 in Rochester, New York.

REFERENCES

American National Standards Institute. (1977). *MARC format for bibliographic data.* Washington, DC: ANSI.

American Library Association. Committee on Cataloging, Description and Access.

(1984). *Guidelines for using AACR2, chapter 9 for cataloging microcomputer software.* Chicago, IL: American Library Association.

Attig, J. C. (1983). The concept of a MARC format. *Information Technology and Libraries, 2*(1), 7-17.

Attig, J. C. (1989). Descriptive cataloging rules and machine-readable record structures: Some directions for parallel development. In E. Svenonius (Ed.), *The conceptual foundations of descriptive cataloging* (pp. 135-148). San Diego, CA: Academic Press.

Baker, P. R. (1985). Adoption of a computer software cataloging system at an elementary school. *School Library Media Quarterly, 13*(3,4), 208-214.

Bates, M. J. (1989). Rethinking subject cataloging in the online environment. *Library Resources & Technical Services, 33*(4), 400-412.

Dumlao, M., & Cook, S. (1983). Cataloging software. *Special Libraries, 74*(3), 240-245.

Gorman, M., & Winkler, P. W. (Eds.). (1988). *Anglo-American cataloguing rules* (rev. 2d ed.). Chicago, IL: American Library Association.

Gorman, M. (Ed.). (1987). *Anglo-American cataloguing rules, second edition. Chapter 9, computer files* (draft revision). Chicago, IL: American Library Association.

Guidelines on subject access to microcomputer software. (1986). Chicago, IL: American Library Association, RTSD/CCS Ad Hoc Subcommittee on Subject Access to Microcomputer Software.

ISBD (CF), International Standard Bibliographic Description, Computer Files. (1989). The Hague: Universal Bibliographic Control Programme.

Koenig, M. E. D. (1990). Linking library users: A culture change in librarianship. *American Libraries, 21*(9), 844-847.

Leysen, J. (1986a). Cutter numbers and subject headings for computer programming languages. *RTSD Newsletter, 11*(5), 49.

Leysen, J. (1986b). Cutter numbers and subject headings for computers, computer systems, and microprocessors. *RTSD Newsletter, 11*(5), 47-49.

Library of Congress. Office of Subject Cataloging Policy . (1990). *Subject cataloging manual: Subject headings* (2d ed.). Washington, DC: Cataloging Distribution Service, Library of Congress.

Library of Congress. Automated Systems Office. (1980). *MARC formats for bibliographic data.* Washington, DC: LC.

Mead-Donaldson, R. (1984). Cataloging microcomputer software at Florida International University. *OLAC Newsletter, 5*(September), 19.

Additional References

American Library Association. Resources and Technical Services Division Audiovisual Committee. (1986). *New directions in subject access to nonbook materials.* Elkridge, MD: Eastern Audio Associates (2 sound cassette recordings).

Dodd, S. A., & Sandberg-Fox, A. M. (1985). *Cataloging microcomputer files: A manual of interpretation for AACR.* Chicago, IL: American Library Association.

Frost, C. O. (1989). *Media access and organization: A cataloging and reference source guide for nonbook materials.* Englewood, CO: Libraries Unlimited.

Holzberlein, D. (1985-86). Computer software cataloging: Techniques and examples. *Cataloging & Classification Quarterly, 6*(2), 1-83.

Intner, S. S., & Smiraglia, R. P. (Eds.). (1987). *Policy and practice in bibliographic control of nonbook media.* Chicago, IL: American Library Association.

Intner, S. S., & Hannigan, J. A. (Eds.). (1988). *The library microcomputer environment: Management issues.* Phoenix, AZ: Oryx Press.

Intner, S. S. (1984). *Access to media: A guide to integrating and computerizing catalogs.* New York: Neal-Schuman Publishers.

Library Hi-Tech. (1985-). Ann Arbor, MI (quarterly with occasional supplements).

National Information Standards Institute. Committee FF: Computer Software Description. (1990). *Draft American National Standard for Computer Software Description* (August revision). Typescript.

Nollan, R. (1986). *Microcomputer software policies in ARL libraries* (SPEC kit 123). Washington, DC: Association of Research Libraries, Office of Management Studies.

OCLC Micro. (1985-). Dublin, OH (bimonthly. Some issues are accompanied by disks containing utility programs).

OLAC Newsletter. (1980-). Tucson, AZ (official publication of Online Audiovisual Catalogers. Quarterly).

Olson, N. B. (1988). *Cataloging microcomputer software: A manual to accompany AACR2, Chapter 9, Computer Files.* Englewood, CO: Libraries Unlimited.

Olson, N. B. (1988). *A manual of AACR2 examples for microcomputer software with MARC tagging and coding* (3d ed.). Lake Crystal, MN: Soldier Creek Press.

T. H. E. (Technological Horizons in Education) Journal. (1983-). Santa Ana, CA (monthly except July and December).

Weihs, J. R. (1984). *Accessible storage of nonbook materials.* Phoenix, AZ: Oryx Press.