Volume 11,
Number 2

ISSN 0730-9295

Information Technology and Libraries

June 1992

CONTENTS

11(2) 89–200 (1992); Inf Tech Libr

ISSN 0730-9295

Information Technology and Libraries

Volume 11, Number 2: June 1992

CONTENTS

ITLBDC 11(2) 89–200 (1992); Inf Tech 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 88 of the March 1992 issue. Manuscripts of articles, communications, and news items should be addressed to the editor: Thomas W. Leonhardt, *Information Technology and Libraries*, University of the Pacific, Stockton, CA 95211. Copies of books for review should be addressed to: Susan B. Harrison, *ITAL Book Reviews*, The New York Public Library, 455 Fifth Ave., New York, NY 10016. Copies of software for review should be addressed to: George S. Machovec, *ITAL Software Reviews*, Arizona State University, Tempe, AZ 85287. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

As a matter of policy, *Information Technology and Libraries*, as the scholarly organ of LITA, does not review LITA publications. Notice of new publications from LITA will generally be found in the "Other Recent Receipts" column following reviews.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $20, is included in membership dues. Nonmembers may subscribe for $40 per year in the U.S.; $50 in Canada, Mexico, Spain, and other PUAS countries; $55 in other foreign countries. Single copies, $15. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. *Postmaster:* Send address changes to Information Technology and Libraries, 50 E. Huron St., Chicago, IL 60611.

Circulation and Production: ALA Publishing Services (Eileen Mahoney; Dianne Rooney; Amy Brown, Bruce Frausto, Josephine Gibson-Porter, Dan Lewis, and Donavan Vicha), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by the Library and Information Technology Association or the American Library Association.

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

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences–Permanence of Paper for Printed Library Materials, ANSI 239.48-1984. ∞

The Need for Machine-Readable Authority Records for Topical Subdivisions

Karen M. Drabenstott

This paper recognizes the limitations of the existing file of Library of Congress subject authority records for subject heading assignment and validation. It makes recommendations for a new machine-readable file of authority records for topical subdivisions and for enhancements to the existing subject authority file. The recommended changes would enable online systems to assist in subject heading formulation and verify, with limited assistance by human intermediaries, the individual components of subdivided headings. A study of subdivided subject headings in a large bibliographic database forms the basis of the recommendations.

No comprehensive list of topical subdivisions is yet available in machine-readable form. The machine-readable *Library of Congress Subject Headings* (*LCSH*-mr) contain a few records for subdivisions when subdivisions are the same as main headings or see references. *LCSH*-mr also contain subdivided headings, but the subdivisions in these headings are only authorized with the particular main headings to which they are appended. Generally, catalogers consult the printed publication *Subject Cataloging Manual: Subject Headings* (*SCM:SH*) to find appropriate subdivisions to append to subject headings.[1]

The availability of machine-readable records for topical subdivisions would enable catalogers to "cut" subdivisions from authority records and "paste" them into bibliographic records. In online cataloging systems, such a capability would reduce typographical errors and minimize the assignment of unauthorized subdivisions. It would not, however, enable systems to determine automatically whether the individual components of subdivided headings are correctly formulated and authorized with the particular main heading. Additional information needs to be incorporated into machine-readable subdivision records to enable systems to perform automatic verification of subdivided subject headings.

The purpose of this paper is to demonstrate the need for machine-readable authority records for topical subdivisions to improve the quality and accuracy of subdivision assignment. It describes the extent to which assigned subject headings in a large bibliographic database are subdivided by topical subdivisions and the source of those subdivisions in printed and online cataloging tools. It makes recommendations for machine-readable authority records for subdivisions and for enhancements to existing files of subject authority records. Such enhancements would enable online systems to verify automatically whether the individual components of subdivided headings are correctly formulated and authorized with the particular main heading.

PREVIOUS CALLS FOR A SUBDIVISIONS FILE

The most widely used subject authority file for subjects is the *LCSH*-mr. Since early 1986, the Library of Congress' Cataloging Distribution Service (CDS) has made *LCSH*-mr avail-

Karen M. Drabenstott is Associate Professor, School of Information and Library Studies, University of Michigan, Ann Arbor.

able to subscribers in the form of a cumulative master tape and a weekly update service. Only one-third of the subject headings in this file are subdivided.[2] In contrast, about two-thirds of the assigned subject headings in bibliographic databases are subdivided.[3,4] The existing subject authority file is of limited use for assignment and validation of subject headings because of the many options available to catalogers for adding subdivisions to the headings printed in *LCSH*.[5]

Since the early 1980s, the library community has called for machine-readable authority files to aid in subject heading assignment and validation. Underlying such calls is indecision about the form of these files. Should these files consist of subdivision records or unique strings of subdivided headings? Such indecision is evident in the recommendation of an ALA subcommittee that encourages the Library of Congress (LC) to conduct research to determine whether separate authority records should be created for every unique heading or whether "separate files of authorized free-floating subject subdivisions would suffice."[6] Pauline Cochrane urges LC to create machine-readable authority records for the subdivisions in the general list of free-floating subdivisions to control the many different misspellings that occur in these frequently used subdivisions.[7] Lois Chan identifies the need for a machine-readable file of unique headings in LC MARC records to assist in the validation of subject entries assigned by cataloging agencies outside LC and to aid in online retrieval.[8]

The Library of Congress Subject Subdivisions Conference held May 9–12, 1991, generated six recommendations. The second recommendation suggests the need for a file of unique strings of subdivided headings.[9] Since the working conference was held so recently, it is too early to determine how LC will proceed. The data and analyses of the study described in this paper could aid LC's decision-making process regarding the form of machine-readable authority files for subject heading assignment and validation.

Online systems provide the "opportunity to detect errors before they are actually entered into a database . . . by incorporating the validation routines into the input/edit process." [10] In online cataloging systems, machine-readable subdivision records could be used to detect errors automatically and correct them with or without the assistance of human intermediaries. Several researchers have studied errors in assigned subject headings with the objective of categorizing them and determining automatic error detection and correction procedures to minimize error occurrences.[11–13] In online bibliographic systems, such procedures could be integrated into a more versatile subject heading validation capability that would not only check for errors but determine whether the individual elements of subdivided subject headings are authorized for use with one another.

RESEARCH QUESTIONS AND METHODS

A study of subdivided subject headings in a large bibliographic database forms the basis of this paper's recommendations on the contents of machine-readable authority records for subdivisions. Three research questions pertain to topical subject headings and topical subdivisions:

1. To what extent do machine-readable and manual sources contribute subdivisions to bibliographic records?
2. What errors in subdivision assignment are connected with the different machine-readable and manual sources of subdivisions?
3. What enhancements to the authority format are needed to improve subdivision assignment and automatic verification?

The OCLC Online Computer Library Center and University of Michigan staffed the study. OCLC Office of Research staff provided Michigan staff with a 0.1% sample of subdivided assigned subject headings for topical subjects and geographic names (MARC tags 650 and 651, respectively) from the OCLC Online Union Catalog (OLUC). Assigned subject headings in the sample are unique strings consisting of one main heading subfield and one or more subfields for subject subdivisions. Data provided by OCLC include number of postings per unique string to enable the Michigan team to count occurrences in addition to unique topical subject headings. Project staff studied topical subject headings separately from geographic subject headings. Staff categorized erroneous main headings, but they treated such headings as if the errors had been corrected. Unauthorized main headings, e.g., obsolete or made-up headings, were omitted from subsequent

analyses because an intellectual decision would be required to replace them with authorized headings before determining what subdivisions were authorized with the newly assigned main headings. The first or main heading subfield of topical subject headings contains subject headings printed in *LCSH*. For this first subfield, Michigan staff determined whether subdivisions from a particular pattern list or free-floating list for groups of headings would apply (i.e., subdivision lists numbered H1100–H1200 in *SCM:SH*). Staff then determined the source of topical subdivisions appended to the main heading subfield by checking the following four sources in this order: (1) subdivided headings in the thirteenth edition of *LCSH*,[14] (2) pattern lists in the third edition of *SCM:SH*,[15] (3) free-floating lists for groups of headings in *SCM:SH*,[16] and (4) the general list of form and topical subdivisions in *SCM:SH*.[17] Staff also generated categories for the erroneous and unauthorized subdivisions that they encountered with a view to making recommendations about the contents of machine-readable subdivision records that would reduce occurrences of such errors in the future.

MAIN HEADINGS FOR TOPICAL SUBJECTS

Categories of Main Headings

Table 1 lists categories of main headings for topical subjects from the 0.1% sample of OLUC headings. Whether the main heading is governed by a particular pattern list or whether a free-floating list for groups of headings can be applied determines the category. Statistics for unique headings are listed in columns on the left; statistics for heading occurrences in the OLUC are listed in columns on the right.

Over 21% of headings for topical subjects are governed by a particular pattern list in *SCM:SH*, i.e., H1145.5–H1200, or are actual pattern headings themselves, e.g., Corn, Cattle, Fishes. Free-floating lists for Classes of persons and Ethnic groups account for 7.8% and 1.5% of main headings for topical subjects, respectively.

Only 0.6% of headings for topical subjects are not explicitly listed in *LCSH*. These are unprinted music headings (0.3%), unprinted chemical headings (0.1%), multiple headings governed by a pattern (0.1%), and headings authorized by a scope note in *LCSH* (0.1%). Examples of subject headings from the multiple headings category are "Theological Seminaries, Baptist" and "Theological Seminaries, Lutheran." These multiple headings fit the pattern for Types of Educational Institutions (H1151.5). Multiple headings have not been treated as free-floating phrases since 1979.[18] They could be considered unauthorized headings; however, few (0.1%) were encountered in the sample.

The two headings authorized by *LCSH* scope notes are "Campaign literature, 1835" and "Songs, Mende." The scope note under "Campaign literature" allows catalogers to qualify the heading by date and subdivide by party and place. The scope note under

Table 1. Categories of Topical Subject Headings

Heading Category	No.	%	No. of Occs.	% of Occs.
Heading fitting a pattern*	618	21.3	3,049	22.8
Classes of persons heading	227	7.8	2,869	21.5
Unauthorized heading	110	3.8	149	1.1
Ethnic group heading	44	1.5	89	0.7
Unprinted music heading†	8	0.3	9	0.1
Unprinted chemical heading†	4	0.1	4	0.0
Heading authorized by an *LCSH* scope note†	2	0.1	3	0.0
Other authorized heading	1,890	65.1	7,190	53.8
Total	2,903	100.0	13,362	100.0

*Includes two multiple headings governed by a pattern that are not printed in *LCSH*.

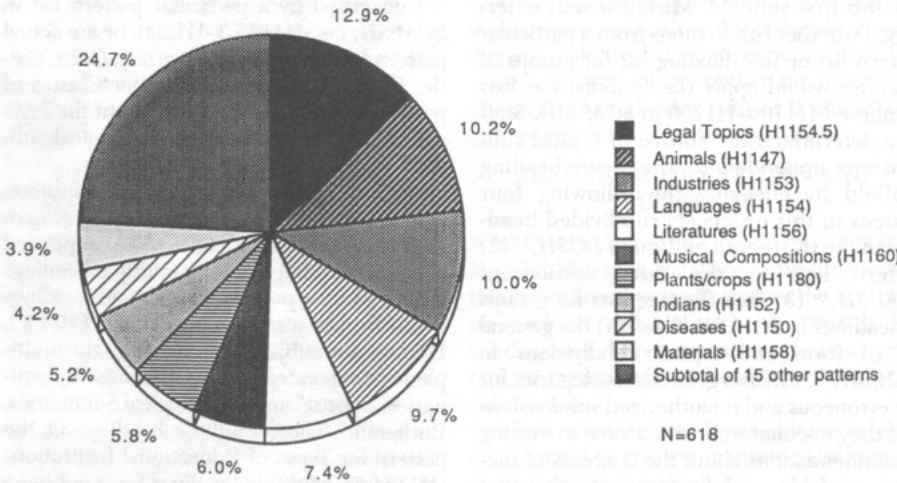


Figure 1. Common Pattern Headings.

"Songs" instructs catalogers to make a subject entry under "Songs" qualified by the "name of the language or language group and subdivided by place, e.g. Songs, French; Songs, Slavic."

SCM:SH authorizes catalogers to add the free-floating phrases "in art" and "in literature" to headings for topical subjects.[19] The sample contained no such unprinted headings.

The Michigan team placed main headings for topical subjects into the last category listed in table 1 for "Other authorized heading" when headings were authorized by the thirteenth edition of *LCSH*, but they failed to qualify for inclusion in any of the other categories. Almost two-thirds (65.1%) of main headings for topical subjects are placed in this category. Examples are "Biculturalism," "Planning," and "Wiretapping." Subdivisions authorized with these main headings are: (1) subdivisions explicitly listed in *LCSH* with the particular main heading, (2) form and topical subdivisions from the general free-floating list (H1095) in *SCM:SH*, and (3) geographic subdivisions for those main headings accompanied by the designation "May subd geog" in *LCSH*.

Main Headings Governed by Pattern Headings

A pattern heading governs a total of 618 main headings (21.3%). Figure 1 shows the ten most common pattern headings that govern main headings in the sample.

Twenty-five different pattern headings govern the sample's main headings for topical subjects. Ten patterns account for over three-quarters of these main headings. The most common pattern is for Legal topics (H1154.5); Animals (H1147), Industries (H1153), and Languages (H1154) follow close behind in terms of percentages of unique main headings.

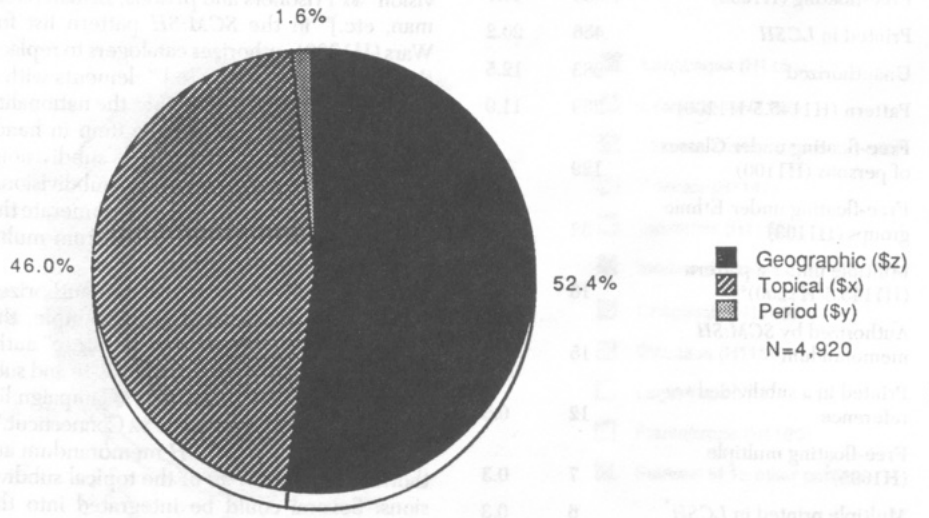


Figure 2. Types of Subdivisions.

TOPICAL SUBDIVISIONS

Types of Subdivisions Appended to Topical Subject Headings

Figure 2 gives the numbers and percentages of the three types of subdivisions appended to 2,903 topical subject headings. The average number of subdivisions per topical subject heading is 1.7.

The majority (52.4%) of subdivisions are geographic subdivisions (i.e., subfield $z). Only 1.6% of the subdivisions appended to headings for topical subjects are period subdivisions. Topical subdivisions account for a large percentage of subdivisions (46.0%). The Michigan team reviewed topical subdivisions to determine how many are form subdivisions. Most form subdivisions, e.g., "Amateur's manuals," "Bibliography," "Encyclopedias," "Maps," and "Software," are easy to spot. If a subdivision could be justifiably classified as both a form and topical subdivision, the Michigan team would choose the latter. Of topical subdivisions, an estimated 53% are form subdivisions. This percentage is comparable to an estimate given in an earlier study.[20] Currently, the USMARC authority format provides no special coding to distinguish topical from form subdivisions.

Overview of Topical Subdivisions

Sources of topical subdivisions are given in table 2. Although the *SCM:SH* list of general subdivisions, i.e., H1095, is the last source checked by the project team, this list contributes the largest percentage (45.7%) of topical subdivisions to main headings for topical subjects. The first source checked, *LCSH*, contributes the second-largest percentage (20.2%) of topical subdivisions. Almost 12% of topical subdivisions are authorized by *SCM:SH* pattern lists.

The free-floating lists for Classes of persons (H1100) and Ethnic groups (H1103) contribute 5.7% and 1.4% of topical subdivisions, respectively. All subdivisions from the free-floating list for the Classes of persons are appended to main headings naming a class of person. Similarly, all subdivisions from the free-floating list for Ethnic groups are

Table 2. Categories of Topical Subdivisions

Subdivision Category	No.	%
Free-floating (H1095)	1,035	45.7
Printed in *LCSH*	456	20.2
Unauthorized	283	12.5
Pattern (H1145.5-H1200)	269	11.9
Free-floating under Classes of persons (H1100)	129	5.7
Free-floating under Ethnic groups (H1103)	32	1.4
Multiple under a pattern (H1145.5-H1200)*	16	0.7
Authorized by *SCM:SH* memorandum*	15	0.7
Printed in a subdivided *see* reference	12	0.5
Free-floating multiple (H1095)*	7	0.3
Multiple printed in *LCSH**	6	0.3
Authorized by an *LCSH* scope note*	3	0.1
Total	2,263	100.0

*Subdivisions not printed in *LCSH* or *SCM:SH*.

appended to main headings naming an ethnic group.

Twelve subdivisions (0.5%) make up subdivided *see* references in *LCSH*. When an authorized heading is substituted for the *see* reference, the authorized heading is also subdivided. An example is the subdivided *see* reference "Chemicals Environmental aspects." The authorized heading for this *see* reference is also subdivided, i.e., *see* "Pollution-Environmental aspects."

Authorized subdivisions not explicitly listed in *LCSH* or *SCM:SH* lists come from three sources: (1) subdivisions created from multiple subdivisions (1.3%), (2) subdivisions authorized by *LCSH* scope notes (0.1%), and (3) subdivisions authorized by *SCM:SH* memoranda (0.7%). Thus, 2.1% of topical subdivisions are not explicitly enumerated in *LCSH* or *SCM:SH* lists.

Multiple subdivisions are authorized through subdivisions printed in *LCSH*, *SCM:SH* pattern lists, and the general list of free-floating subdivisions in *SCM:SH*. Multiple subdivisions save space in printed sources because all possible combinations do not have to be listed. For example, the multiple subdivision "$x Prisoners and prisons, British [German, etc.]" in the *SCM:SH* pattern list for Wars (H1200) authorizes catalogers to replace the "British [German, etc.]" elements with a proper adjective that describes the nationality of the group discussed in the item in hand. The three categories of multiple subdivisions in table 2 total 1.3% of topical subdivisions. Print and online sources do not enumerate the unique subdivisions constructed from multiple subdivisions.

Three subdivisions (0.1%) are authorized by an *LCSH* scope note. For example, the scope note under "Campaign literature" authorizes the cataloger to qualify by date and subdivide by party and place, e.g., "Campaign literature, 1835 $x Democratic $z Connecticut."

The text of an *SCM:SH* memorandum authorizes fifteen (0.7%) of the topical subdivisions. Several could be integrated into the general list of free-floating subdivisions accompanied by footnotes restricting their use.

Subdivisions Authorized by *SCM:SH* Pattern Lists

An *SCM:SH* pattern list authorizes 11.9% of the topical subdivisions. The 0.1% sample yielded subdivisions from twenty-five different patterns. Figure 3 gives the ten most frequently occurring pattern lists.

Although figures 1 and 3 refer to pattern lists, percentages are not the same. Figure 1 gives percentages of main headings governed by specific pattern lists, whereas figure 3 gives percentages of subdivisions actually derived from the pattern lists. Both tables contain only one pattern list that is unique to each (figure 1, Materials; figure 3, Chemicals).

In figure 3, subdivisions from the Languages pattern list (H1154) are 23.8% of topical subdivisions from pattern lists. Pattern lists for Literatures (H1156) and Musical Compositions (H1160) also contribute large percentages of subdivisions. Pattern lists for Christian Denominations (H1187) and Land Vehicles (H1195) did not contribute subdivisions to the sample of topical subdivisions; however, some main headings for topical subjects do fit these patterns.

Authority records for topical subject headings should be enhanced with a designation for the particular *SCM:SH* subdivision list(s)

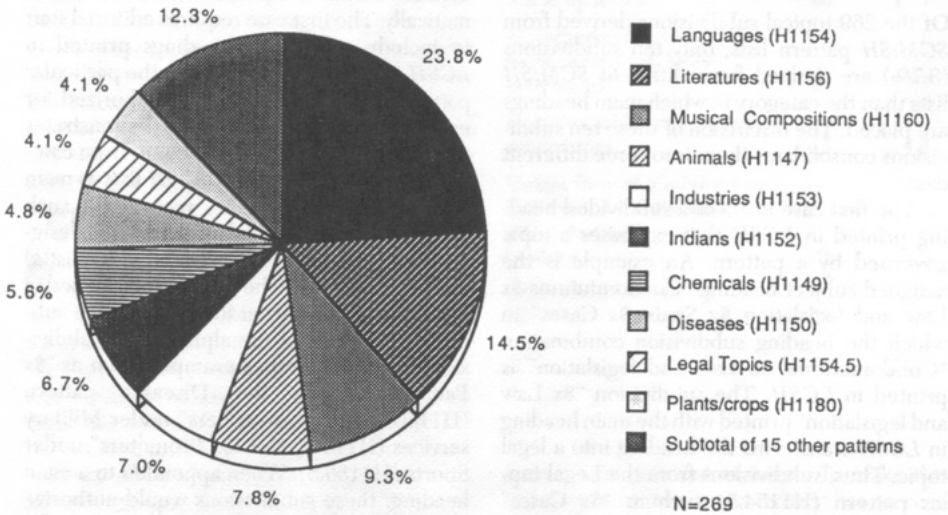


Figure 3. Subdivisions from SCM:SH *Pattern Lists.*

bearing subdivisions authorized for use. This designation need not include the general free-floating list, i.e., H1095, because subdivisions from this list are appropriate for use under subject headings generally, i.e., MARC tags 6xx.

Machine-readable authority records for subdivisions should also designate the particular *SCM:SH* subdivision list(s) from which they are derived. Online systems can then be programmed to check the designations on authority records for both main headings and subdivisions to determine whether a particular combination is valid.

An example illustrates how verification can be done automatically. In the assigned subject heading "650 Helminthiasis $x Diagnosis," the main heading "Helminthiasis" describes a disease. Its authority record would contain the designation H1150 to indicate that it is governed by the Diseases pattern. Thus, subdivisions from the Diseases pattern, i.e., H1150, would be authorized for use with this particular main heading. The authority record for the subdivision "$x Diagnosis" would contain one designation for the Diseases pattern, i.e., H1150, to indicate that this subdivision resides only on the Diseases pattern list. When the cataloger adds a bibliographic record to a cataloging database bearing the assigned heading "Helminthiasis $x Diagnosis," the online system would validate the main heading through the subject authority record for the main heading "Helminthiasis" and the subdivision through the subject subdivision authority record for the subdivision "$x Diagnosis." Since both authority records contain the same designation H1150, the system

automatically determines that the combination of main heading and subdivision is valid.

Different Pattern Lists for Main Headings and Subdivisions

Of the 269 topical subdivisions derived from *SCM:SH* pattern lists, only ten subdivisions (3.7%) are derived from different *SCM:SH* lists than the category in which main headings are placed. The discussion of these ten subdivisions consolidates them into three different cases.

The first case involves a subdivided heading printed in *LCSH* that expresses a topic governed by a pattern. An example is the assigned subject heading "Condominiums $x Law and legislation $z Spain $x Cases" in which the heading subdivision combination "Condominiums $x Law and legislation" is printed in *LCSH*. The subdivision "$x Law and legislation" printed with the main heading in *LCSH* transforms the heading into a legal topic. Thus, subdivisions from the Legal topics pattern (H1154.5) such as "$x Cases" would be authorized with the preceding elements of the string.

The second case involves a subdivision from one pattern that is appended to a main heading subdivision combination from a different pattern. The assigned subject heading is "Pineal gland $x Diseases $x Immunological effects." Printed in *LCSH*, the main heading is governed by the Organs/regions of the body pattern (H1164). The first-listed subdivision comes from the same pattern. It transforms the heading into a disease topic, and thus, subdivisions from the Disease pattern such as "$x Immunological effects" are now authorized with the preceding two elements.

The third case involves a pattern authorized by a main heading subdivision combination. When the subdivision "$x Colonies" is used under "names of countries," it transforms the main heading subdivision(s) combination into a colonies topic. Subdivisions from the Colonies pattern (H1149.5) are then authorized with preceding elements of the string. An example is the assigned subject heading "Slavery $z France $x Colonies $x History." The main heading "Slavery" comes from the "Other authorized heading" category. When the place name "$z France" is appended to this main heading, it authorizes the addition of the subdivision "$x Colonies" from the Colonies pattern (H1149.5) and transforms the string into a colonies topic. The last subdivision comes from the Colonies pattern (H1149.5).

Different actions must be taken to enable online systems to recognize each case automatically. The first case requires editorial staff to include subdivided headings printed in *LCSH* in a review to determine the particular pattern and free-floating lists authorized for use with them. The second case demonstrates that subdivisions sometimes come from categories naming different *SCM:SH* lists as main headings. When authority records for such subdivisions are created, they should designate the particular *SCM:SH* subdivision list(s) authorized for use. The examples discussed in this paper are a start at identifying such subdivisions. Scanning the alphabetical subdivision index yields other examples such as "$x Patients" under the Diseases pattern (H1150), "$x Patternmakers" under Military services (H1159), and "$x Promoters" under Sports (H1188.5). When appended to a main heading, these subdivisions would authorize subdivisions from the *SCM:SH* list for "Classes of persons" (H1100). Determining valid lists is not going to be readily apparent by scanning subdivisions. Often, subdivisions have to be used with a main heading to demonstrate how their use transforms the meaning of the string. The third case is rather difficult to handle because it involves authorization of topical subdivisions from a pattern for specific heading subdivision combinations. Instead of programming systems for individual cases, system designers could allow catalogers to override a system prompt, warning them that they may have made an incorrect heading subdivision combination.

To improve automatic verification, system designers should give priority to the following: (1) enhancing authority records for topical subject headings with a designation for the particular *SCM:SH* subdivision list(s) bearing subdivisions appropriate for use with the particular main heading, and (2) creating authority records for subdivisions that designate the particular *SCM:SH* subdivision list(s) from which they are derived.

Topical Subdivisions Authorized by *SCM:SH* Memoranda

Fifteen (0.7%) topical subdivisions are authorized by the text of an *SCM:SH* memorandum. One such subdivision is "States," which

occurs only on the free-floating list under places (H1140). The text of *SCM:SH* memorandum H713 authorizes this subdivision under first-order political divisions naming countries. Thus, this topical subdivision can be an element in a string bearing a main heading for topical subjects (MARC tag 650) and geographic subdivision (subfield code $z). Examples of assigned subject headings in which the use of this subdivision is authorized are "Day care centers $z United States $x States" and "Representative government and representation $z United States $x States."

The sixth recommendation of the LC Subject Subdivisions Conference addresses simplification of the *LCSH* subdivision system.[21] The topical subdivisions in this category are candidates for simplification. For example, they could be consolidated in the general free-floating list, and overly fine distinctions regarding use could be eliminated. Until LC undertakes a simplification project, online systems could report such subdivisions to human intermediaries who would have to handle each one on an individual, case-by-case basis.

Unauthorized Topical Subdivisions

Of the topical subdivisions appended to headings for topical subjects, 12.5% are unauthorized (see table 2). Table 3 lists categories of unauthorized subdivisions.

The majority (56.9%) of unauthorized subdivisions are obsolete subdivisions, e.g., "$x Addresses, essays, lectures," "$x Yearbooks," that LC canceled in recent years. The availability of subdivision authority records for obsolete subdivisions would assist systems staff who are working with a file of cataloging records created over a long period of time to identify and delete such subdivisions. In the sample, a handful of obsolete subdivisions contain errors, such as transposed letters or additional words, that would have to be corrected before subdivision authority records could detect and delete them. One-eighth of unauthorized subdivisions are made-up terms, for example, $x Auditing and inspection, $x Book lists, $x Collections, $x Production management, $x Recordings, $x Securities, $x (Selections : Extracts, etc.), and $x Theses.

Another eighth of unauthorized subdivisions are unauthorized with the particular main headings to which they are appended because they are out of scope. To make this decision, a human intermediary has to read the *SCM:SH* scope note for the subdivision and evaluate the use of the particular subdivision in the context of the entire string. An example is "Historic sites $x Preservation." The topical subdivision in this subject heading is unauthorized because the scope note in the general list (H1095) instructs catalogers to use the topical subdivision "$x Preservation" under "types of perishable products, including food, drugs, textiles, etc." The unauthorized headings placed in this category require a careful reading and evaluation of scope notes in the general free-floating list (H1095).

Table 3. Unauthorized Topical Subdivisions

Notes	No.	%
Obsolete words or phrases	161	56.9
Made-up terms	36	12.7
Unauthorized according to scope notes in *SCM:SH* memoranda	36	12.7
Unauthorized use of topical subdivisions	13	4.6
Variant form of a subdivision that would be authorized with a particular main heading	9	3.2
Other unauthorized subdivisions	28	9.9
Total	283	100.0

Unauthorized use of topical subdivisions accounts for 4.6% of unauthorized subdivisions. An example is the assigned subject heading "Thunder $z Oklahoma $x Rites and ceremonies." The topical subdivision "$x Rites and ceremonies" is authorized for use under Ethnic groups (H1103); however, "Thunder" is not an ethnic group. Online systems could detect automatically subdivisions placed in this category using these data in authority records: (1) main heading records with coding as to the *SCM:SH* subdivision lists that are appropriate for use with them, and (2) subdivision records with coding as to the *SCM:SH* lists from which they are derived. Systems could not, however, read scope notes and footnotes to make subtle distinctions about the use of subdivisions with restrictions.

Less than a dozen subdivisions look like made-up terms but they are variant forms of topical subdivisions. For example, the unauthorized subdivision "$x Accounts" is a variant form of the topical subdivision "$x Accounting" from the general free-floating list.

Nine separate categories have been collapsed into the "Other unauthorized subdivisions" category in table 3. This one category accounts for nearly one tenth of unauthorized subdivisions. A description of the subcategories making up this category is interesting. Some subdivisions are unauthorized for use as subdivisions but they are authorized for use as main headings. A few unauthorized subdivisions are composed of words and phrases from authorized main headings.

Some unauthorized subdivisions that are words or phrases that are part of main headings could be detected automatically using the existing file of subject authority records. When no authority record exists for the main heading, systems could look for the combination of main heading and first-listed topical subdivision without the presence of the subfield code. For example, the system encounters the main heading subdivision combination "Prison $x Furloughs." No machine-readable record exists for the main heading "Prison," so the system composes the main heading "Prison furloughs." The system validates the newly composed heading with the authority record for the main heading "Prison furloughs."

A few unauthorized subdivisions occur in subdivided headings with an *authorized* main heading. Automatic detection and correction of these unauthorized subdivisions would be facilitated by the availability of machine-readable subdivision records. For example, the system encounters the main heading subdivision combination "Glass $x Optical." An authority record exists for the main heading "Glass," but not for the topical subdivision "$x Optical." The system composes the main heading "Glass optical" which is validated with the normalized form of the authorized heading "Glass, Optical." Other unauthorized subdivisions are split *see* references in *SCM:SH*, corporate names, exact multiple subdivision (including sample terms and brackets), and *see* references from *LCSH* or *SCM:SH* free-floating lists.

Table 4. Erroneous Topical Subdivisions

Type of Error	No.	%
Main heading and subdivision make up a *see* reference	21	27.2
Addition/omission of one character	16	20.8
Incorrect subfield code	15	19.5
Abbreviation error	8	10.4
Punctuation error	6	7.8
Substitution/transposition error	6	7.8
See reference *SCM:SH*	3	3.9
Spelling error	2	2.6
Total	77	100.0

Erroneous Topical Subdivisions

Michigan staff categorized erroneous subdivisions prior to further analysis of the 0.1% sample. Of topical subdivisions appended to headings for topical subjects, 3.2% have one error and 0.2% have two errors. Table 4 lists errors occurring in topical subdivisions.

The largest category of errors is for main heading subdivision combinations that make up a *see* reference (27.2% of errors). These *see* references could be replaced with authorized headings using the *see* references in *LCSH*-mr. The availability of machine-readable records for subdivisions would make it possible for online systems to detect and correct incorrect subfield codes (19.5% of errors) and punctuation errors (7.8% of errors), and to replace occurrences of *see* references for subdivisions with authorized subdivisions (3.9% of errors).

Remaining error categories describe typographical errors and account for 41.6% of errors. Errors in these categories would have to be corrected with the aid of a human intermediary in view of difficulties researchers have encountered using automatic spelling detection or correction algorithms.[22] For example, correction of the main heading "Fusoin" requires a human intermediary to determine which heading, "Fusain" or "Fusoin," is appropriate to assign to particular bibliographic record at hand.

Although only two subdivisions contain more than one error, they would be difficult to detect automatically. For example, the subdivision "$y Early works tp 1800" contains a subfield coding error and substitution error.

The existence of the latter would make it difficult for the system to detect the former.

MACHINE-READABLE AUTHORITY RECORDS FOR SUBDIVISIONS

Existing authority records in *LCSH*-mr assist catalogers in the assignment and validation of *unsubdivided* subject headings; however, the majority of assigned subject headings in bibliographic databases are *subdivided.* To assist in the assignment and validation of *subdivided* headings, catalogers need machine-readable authority records for subdivisions. This section enumerates and discusses recommendations regarding the content of such records.

1. *Machine-readable authority records for topical subdivisions must include a designation that specifies the particular* SCM:SH *list(s) in which the subdivision resides. SCM:SH* contributes topical subdivisions to assigned subject headings governed by pattern lists (H1145.5-H1200) or by lists of free-floating subdivisions (H1095-H1140). For example, the subdivision "$x Uniforms" resides on the *SCM:SH* free-floating list for Classes of persons (H1100), and on the pattern lists for Military Services (H1159) and Sports (H1188.5). The subdivision authority record for "$x Uniforms" would contain designations for the free-floating (H1100) and pattern (H1159 and H1188.5) lists to facilitate automatic validation.

2. *Machine-readable* LCSH *records must be enhanced with a designation that specifies the particular* SCM:SH *list(s) that are authorized for use with the subject heading in the established heading field.* For example, the topical subject "Basketball" is governed by the pattern list of subdivisions for Sports (H1188.5). Thus, the authority record for the topical subject "650 Basketball" would contain the designation H1188.5 to indicate that catalogers may add subdivisions from the Sports pattern.

Catalogers can use *LCSH*-mr to "cut and paste" the topical subject heading "Basketball" into the assigned subject heading they are creating. Systems can be programmed to make subdivisions from the general list (H1095) and from the Sports pattern (H1188.5) available to catalogers. (If geographic subdivision is authorized for use with this main heading, systems can also prompt catalogers to add geographic subdivisions to the assigned subject heading they are constructing.)

A different approach would allow catalogers to append subdivisions from any *SCM:SH* list to the assigned subject heading they are constructing. After they finish the heading, online systems would automatically validate the individual elements of the string and the entire subdivided heading and, if needed, ask catalogers for assistance. For example, if *LCSH*-mr and subdivision records were enhanced with *SCM:SH* designations, it would be possible for online systems to validate automatically the heading-topical subdivision combination "650 Basketball $x Uniforms" without catalogers' assistance, because the authority records would contain the same designation, i.e., H1188.5.

3. LCSH-*mr records for subdivided headings also require designations that specify the particular* SCM:SH *list(s) that catalogers may use.* Some heading subdivision combinations transform the string into a heading governed by subdivisions from different patterns or lists than the main heading alone—for example, "Condominiums $x Law and legislation." Systems would be able to detect subdivisions added to subject heading strings that are derived from unauthorized lists. They would not, however, be able to detect nonsensical subdivided headings because they cannot read or interpret the meaning of subject headings.

4. *Subdivisions on* SCM:SH *pattern and free-floating lists require review to determine whether their use in subject heading strings changes the meaning of strings, and thus, authorizes the use of subdivisions from a different pattern or free-floating list.* Reviewing subdivisions on *SCM:SH* pattern and free-floating lists to determine whether their use in subject heading strings changes the meaning of strings, and thus authorizes the use of subdivisions from a different pattern or free-floating list, will not be an easy task. Reviewers will have to scrutinize some subdivisions in the context of subdivided headings to determine if they change meaning. However, some subdivisions such as those naming classes of persons, e.g., "$x Parachute troops," "$x Patients," or "$x Paternmakers," and those naming actual patterns themselves, e.g., "$x Colonies," "$x Diseases," and "Languages," are easy to spot in the alphabetical list of free-floating subdivisions.[23]

5. *Adding* SCM:SH *scope notes, foot-*

notes, and see *references would improve machine-readable subdivision records.* The presence of scope notes and footnotes in machine-readable authority records for subdivisions could reduce unauthorized uses of subdivisions, because these notes would be available to catalogers on the authority records from which they "cut" subdivisions to "paste" into the assigned subject heading they are creating. Of unauthorized topical subdivisions, 12.7% are used with headings for topical subjects that are out of scope (table 3). In the future, fewer scope notes and footnotes associated with subdivisions may be used in view of the sixth recommendation of the Library of Congress Subject Subdivisions Conference that calls for reduction of overly fine distinctions in the use of subdivisions. The decision to use a subdivision with a particular heading may become increasingly dependent on the cataloger's good judgment, rather than on specific rules, instructions, and notes in *SCM:SH* memoranda and in machine-readable authority records. Few unauthorized subdivisions are *see* references, but the availability of *see* references in authority records for subdivisions would enable systems to substitute automatically the authorized subdivision in place of the *see* reference.

6. *Subdivisions discussed exclusively in* SCM:SH *memoranda should be integrated into the general free-floating list accompanied by scope notes restricting use.* The researchers encountered few occurrences of subdivisions that are governed by instructions in *SCM:SH* memoranda. Online systems can be programmed to handle these types of subdivisions much as they handle subdivided subject headings bearing subdivisions from different pattern or free-floating lists. (This assumes that the recommended coding for pattern and free-floating lists is added to authority records for *LCSH*-mr and subject subdivisions.) When a special subdivision appears in a heading, the system could alert the catalogers, who could then review the heading to determine whether the main heading-subdivision combination is authorized or needs to be changed.

The sixth recommendation of the Library of Congress Subject Subdivisions Conference urges LC to simplify the *LCSH* subject subdivisions system. If LC addresses the two areas of simplification given in the recommendation, i.e., reduction of overly fine distinctions and consolidation of lists, the number of subdivisions whose use is restricted by instructions in *SCM:SH* could be considerably reduced if not eliminated altogether. For example, the topical subdivision "$x Religious aspects" could be mainstreamed into the general list (H1095) with a note authorizing use under nonreligious topics that are addressed from a religious viewpoint.

7. *Catalogers need machine-readable records for obsolete and canceled subdivisions to facilitate their deletion from older cataloging records.* If library staff are working with an old file of cataloging records, they can expect occurrences of obsolete and canceled subdivisions, e.g., "$x Addresses, essays, lectures" and "Yearbooks." Cross-references should be made for errors that occur frequently with obsolete subdivisions, e.g., incorrect subfield codes or transposed letters.

DISCUSSION

The data and analyses in this paper demonstrate the potential of online systems to assist in the assignment and validation of subject headings. To ensure the precision of system validation, the *LCSH*-mr records must contain a designation that specifies the particular *SCM:SH* list(s) that catalogers may use with the subject in the established heading field. Machine-readable authority records for subdivisions must also contain a designation that identifies the particular *SCM:SH* lists from which they are derived. Such designations would enable online systems to determine automatically that the topical subdivisions appended to the main heading are authorized for use.

In the absence of the recommended designations, automatic validation would require authority records for every possible combination of heading and topical subdivision. Creating all possible combinations would be a time-consuming and expensive manual effort. For example, the free-floating list for Classes of persons (H1100) contains about 180 subdivisions.[24] *LCSH*-mr contains about 146,500 authority records for topical subjects.[25] The number of Classes of persons headings is estimated to be 11,500 (7.8% of topical subjects; see table 1). All 11,500 main headings for classes of persons headings can be appended by the 180 subdivisions from this list. Thus, a total of 2.07 million subject authority records

would be required to document all possible combinations of classes of persons headings and subdivisions from the free-floating list for Classes of persons (H1100).

This number is just the tip of the iceberg with respect to creating authority records for every unique heading-subdivision combination. *SCM:SH* contains a half dozen free-floating lists (H1095-H1140) and nearly three dozen pattern lists (H1145.5-H1200). It would probably be quicker to single out *LCSH* headings governed by *SCM:SH* free-floating or pattern lists, create minimal-level authority records for subdivisions from each list, and employ a computer program to generate all possible combinations. Some combinations would not make much sense, e.g., "651 Great Britain $x History $y Wars of the Roses, 1455 -1485 $x Tank warfare" or "650 Toes $x Left ventricle." However, automatic generation of all possible combinations would be much faster than waiting for LC (or any library or group of libraries) to construct authority records manually for use in automatic validation.

The library community's interest in a file of unique heading subdivision combinations is manifest in the second recommendation of the LC Subject Subdivisions Conference, which describes the creation of a "national authority file" of "authority records for topical headings and for topical heading topical subdivision(s) combinations" and "headings containing subdivisions governed by pattern and free-floating lists."[26] Before the library community embarks on ambitious projects to create such a file, it should consider the alternative approach using machine-readable files of subdivision records.

SUMMARY

The usefulness of the existing file of subject authority records is limited because it contains mostly unsubdivided headings. In a bibliographic file, most subject headings are subdivided.

This paper bases recommendations for machine-readable authority records on a study of subdivided subject headings in a large bibliographic database. It calls for a new machine-readable file of authority records for topical subdivisions and suggests enhancements to the existing subject authority file to enable online systems to assist in subject heading formulation and to verify, with limited assistance by human intermediaries, the individual components of subdivided headings. The paper also makes recommendations regarding the content of the new and existing files of authority records.

Implementing the recommendations would allow effective systems designs for incorporating machine-readable subdivision records in online systems for formulating and validating assigned subject headings. Library staff and patron users will ultimately benefit from such records because fewer errors and unauthorized headings will occur in catalog access points.

ACKNOWLEDGMENT

The author acknowledges the support of the OCLC Online Computer Library Center for its support of this project. The author also thanks Nancy Sack, presently Library Associate, University of Illinois at Chicago, who categorized the sample of topical subject headings. Bonnie A. Dede, Head, Special Formats, University of Michigan libraries, served as project consultant.

REFERENCES

1. Library of Congress, *Subject Cataloging Manual: Subject Headings*, 3d ed. (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1990).
2. Karen Markey and Diane Vizine-Goetz, *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings*, OCLC Research Report Series OCLC/OR/RR-88/2 (Dublin, Ohio: OCLC, 1988), p.67.
3. Carolyn O. Frost and Bonnie A. Dede, "Subject Heading Compatibility between LCSH

and Catalog Files of a Large Research Library: A Suggested Model for Analysis," *Information Technology and Libraries* 7, no.3:291 (Sept. 1988).
4. Edward T. O'Neill and Rao Aluri, *Subject Heading Patterns in OCLC Monographic Records*, OCLC Research Report Series OCLC/RDD/RR-79/1 (Columbus, Ohio: OCLC, 1979), p.80.
5. Mary K. D. Pietris, "Characteristics of LC Subject Headings that Should be Taken into Account when Designing On-line Catalogues," *Cataloguing Australia* 13, no.3:78-79 (1987).
6. Ad Hoc Subcommittee on Library of Congress Subject Authority Control, "Library of Congress Subject Authority Control: Scope, Format, and Distribution; A Final Report," photocopy submitted to the Subject Analysis Committee of the American Library Association, July 12, 1982.
7. Pauline A. Cochrane, *Improving LCSH for Use in Online Catalogs* (Littleton, Colo.: Libraries Unlimited, 1986), p.59.
8. Lois Mai Chan, *Library of Congress Subject Headings: Principles and Application* (Littleton, Colo.: Libraries Unlimited, 1986), p.360.
9. Library of Congress, *Subject Subdivisions Conference* (Washington, D.C.: Catalog Distribution Service, in press).
10. Jeanne C. Sawyer and Jinnie Y. Davis, "Automated Error Detection in Library Systems," in *Crossroads: Proceedings of the 1st National Conference of the Library and Information Technology Association*, ed. Michael Gorman (Chicago: American Library Assn., 1984), p.217.
11. Edward T. O'Neill, "Database Quality," in *OCLC Annual Review of Research* (Dublin, Ohio: OCLC, 1988), p.13 -14.
12. Anne E. Lantz and others, "'Addrdsses, Elctures, Esssays' and Other Variant Subject Headings," in *Academic Libraries: Myths and Realities: Proceedings of the Association of College and Research Libraries (ACRL) 3rd National Conference*, ed. Suzanne C. Dodson and Gary L. Menges (Chicago: ACLR, 1984), p.183–86.
13. Frost and Dede, "Subject Heading Compatibility," p.294–97.
14. Library of Congress, *Library of Congress Subject Headings*, 13th ed. (Washington, D.C.: Cataloging Distribution Service, 1990).
15. Library of Congress, *Subject Cataloging Manual*, H1145.5-H1200.
16. Ibid., H1100-H1140.
17. Ibid., H1095.
18. Chan, *Library of Congress Subject Headings*, p.174.
19. Library of Congress, *Subject Cataloging Manual*, H362.
20. O'Neill and Aluri, *Subject Heading Patterns*, p.11.
21. Library of Congress, *Subject Subdivisions Conference*, in press.
22. Edward T. O'Neill and Diane Vizine-Goetz, "Quality Control in Online Databases," in *Annual Review of Information Science and Technology*, ed. Martha E. Williams (Amsterdam: Elsevier Science Pubs., 1988), p.11.
23. Library of Congress, *Free-floating Subdivisions: An Alphabetical Index* (Washington, D.C.: Cataloging Distribution Service, 1990).
24. Library of Congress, *Subject Cataloging Manual*, H1100.
25. Library of Congress, *Library of Congress Subject Headings*, 14th ed. (Washington, D.C.: Cataloging Distribution Service, 1991), p.vii.
26. Library of Congress, *Subject Subdivisions Conference*, in press. ■■

Nonroman Scripts in the Bibliographic Environment

Joan M. Aliprand

The representation of nonroman scripts in Latin characters causes information to be distorted in various ways. USMARC now provides for "alternate graphic representation," so that text in the authentic script(s) may be included in bibliographic records. As more library systems with nonroman capability are developed, conformance to standards for the encoding of nonroman data becomes more critical. The development of a single global character set standard is a significant change that must be accommodated in USMARC.

In Rule 1.0E, *AACR2* mandates that the bibliographic description be written in the same script as the source of information "if practicable."[1] For more than a decade, machine-readable cataloging and bibliographic transcription in a nonroman script were mutually exclusive. During this period, the only way to represent nonroman data in machine-readable form was by transcription into Latin letters (romanization). The first part of this paper criticizes romanization as information distortion.

The *USMARC Format for Bibliographic Data* was modified to accommodate nonroman scripts in 1984.[2] The previous September, a Chinese/Japanese/Korean (CJK) capability had been added to the Research Libraries Information Network (RLIN) system.[3] The USMARC modifications are outlined in the second part of this paper, since not all readers will be familiar with them. The remainder of this paper describes efforts to develop a universal character set, and its potential effect on USMARC.

ROMANIZATION AS INFORMATION DISTORTION

Currently, most local systems are limited to Latin script; romanization is necessary if the automated catalog is to be a comprehensive representation of the library's holdings. The practice of romanization has two causes: the lack of the proper typographical facilities and the concept of the "universal" catalog, "the catalog in which all items in the collection are entered in a single alphabet from A to Z, regardless of language, regardless of form, regardless of subject. The American ideal."[4]

The deficiencies of romanization from the point of view of the reader have been documented.[5–7] However, many nonspecialist librarians are unaware of the deficiencies and still regard romanization as adequate for access. Language experts reject this view; they persuaded the Library of Congress (LC) to continue to provide original script cataloging on cards for material in the so-called JACKPHY languages: *J*apanese, *A*rabic, *C*hinese, *K*orean, *P*ersian (Farsi), *H*ebrew, and *Y*iddish.

Not only does romanization impede access, it distorts the presentation of information in a number of ways. The presentation of the text is unnatural. Distinctions present in the original language may be lost, or distinctions not present in the original script may be artificially created. Different transliteration schemes are used in different countries or contexts. Finally, the normalization used in automated indexing and searching, when applied to romanized text, introduces another layer of distortion.

Joan M. Aliprand is a programmer/analyst with The Research Libraries Group, Inc., Mountain View, California. This paper represents the author's own views and not necessarily those of RLG.

Unnatural Presentation

Romanization is the presentation of language text in unfamiliar letters. Readers of a language may, in time, become used to a particular romanization scheme, and be able to read their language even when it is written in Latin letters. In the People's Republic of China, pinyin, the national standard for the romanization of Chinese, has a number of applications: it is used to show the pronunciation of ideographs (in which Chinese is normally written), and it underlies a system of fingerspelling for the blind.

A reader faced with text rendered in an unfamiliar way may find it incomprehensible. This can be illustrated by the case of alternative romanization methods. Hebraica bibliographers in the United States have become used to reading Hebrew written in Library of Congress romanization (which includes the vowels that are usually omitted in Hebrew orthography). An alternative romanization scheme for Hebrew is the American National Standards Institute (ANSI) reversible romanization (which was used in the production of the New York Public Library's *Dictionary Catalog*) (see figure 1).[8] The author has been told by more than one specialist in Hebraica that ANSI reversible romanization is "unreadable," despite the fact that it is closer to the original text, being a simple substitution (usually one to one) of Latin letters for Hebrew letters. However, Malinconico reported, "A reader familiar with a language written in Hebrew script, after a short time for familiarization with the correspondences [of the letters], can read text represented in ANSI transcription with practically no impediment."[9]

The perplexity experienced by these librarians when faced with the unfamiliar, reversibly romanized Hebrew is similar to the reaction of a reader presented with a romanized catalog entry for a work written in a language that he knows. The catalog provides neither a key to the romanization methodology nor a clue to the pronunciation of any accented Latin letters in the romanization.

For example, Chinese people may be accustomed to Chinese rendered with pinyin, but may be confounded by catalog entries for Chinese books romanized according to the Wade-Giles method used by American libraries. The difficulty that Chinese people have interpreting Wade-Giles romanization may be partly due to the fact that the pronunciation of the Wade-Giles markings is not self-evident. When a reader does not know how a diacritical mark affects pronunciation, the diacritical mark is useless. Watts, writing before the advent of pinyin, said, "For Mandarin one is almost compelled by general usage to adopt the Wade-Giles romanization, for which I have appended a table of pronunciation . . . , since it has so little relation to the actual sounds."[10]

The problem is not limited to Chinese. Persian (Farsi) is written in Arabic script. The Library of Congress' romanization tables for the language reflect the Arabic pronunciation of consonants. As a result, romanized Persian has little resemblance to the source language.[11] When Spurrier demonstrated RLIN searching at the Kuwait Institute of Scientific Research in March 1990, the romanized Arabic in bibliographic records did not convey information to readers of Arabic.[12] Romanized Arabic also includes nonalphabetic marks: special characters for the letters *ayn* and *alif*, and diacritical dots.

Loss of Distinctions

Chinese is an example of a language where distinctions in the original script are lost when text is romanized. Chinese contains a considerable number of homophones, and the differences between characters with the same sound cannot be maintained in a romanization scheme based on the rendering of pronunciation. For example, the Chinese words mean-

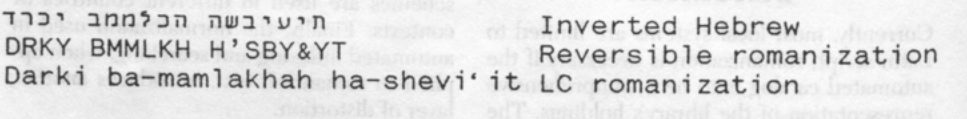


Figure 1. ANSI Reversible Romanization for Hebrew. The illustration shows a phrase written in Latin letters according to ANSI reversible romanization and LC romanization, repectively. (The Hebrew text is written left to right, that is, backwards, to make comparison easier.)

ing "king" and "die" sound exactly the same, but they are written with completely different characters. Since both pinyin and Wade-Giles romanize on the basis of pronunciation, the distinction in the original text cannot be maintained in romanization; both characters in this example are represented by a single Latin letter sequence ("wáng" in pinyin romanization).[13]

Differences in sound are shown by different spellings, augmented with tone marks. Both the pinyin and the Wade-Giles systems reflect Mandarin Chinese pronunciation, and both include mechanisms to indicate its tones: pinyin uses accent marks, and Wade-Giles usually has superscript numbers. For example, when tone marks are used, "wàng," the romanization of the character that means "absurd," is quite distinct from "wáng," which represents either "king" or "die." The distinguishing tone marks are not used in cataloging, thus increasing the number of identical spellings for romanized Chinese (and creating false drops in retrieval).

Pinyin provides for a greater number of discrete letter combinations than the Wade-Giles system; the latter makes some distinctions by applying diacritical marks to otherwise identical spellings. For example, the words ping and p'ing (distinguished only by the aspirate) are spelled as bing and ping in pinyin. The 404 discrete letter combinations of pinyin correspond to only 330 Wade-Giles spellings when diacritical marks and special characters are ignored.

Despite the widespread use of pinyin romanization, American libraries still use the Wade-Giles system to romanize Chinese; one reason is the lack of standard rules for word division in pinyin. Under Wade-Giles romanization, the amalgamation of discrete Chinese characters has three causes: the unavoidable problem of identical pronunciation, the deliberate omission of distinguishing tone marks, and the deliberate dropping of diacritical marks in indexing and searching (which is discussed in more detail below).

Failure to maintain distinctions when text is romanized is not limited to Chinese. Variant orthography is a longstanding feature of the Hebrew language; Yiddish orthography was not standardized until 1936. The Library of Congress romanization obscures some orthographic features of original Hebrew or Yiddish text; in particular, different orthographies may be unified. Weinberg has described in some detail the issues associated with the transliteration of variant Hebrew and Yiddish orthography.[14] When languages usually written without vowels (such as Hebrew and Arabic) are romanized, vowels are always supplied; thus, the orthography of the original text is not represented (and it may sometimes be significant). The presence of aberrant letter forms in Soviet Yiddish orthography cannot be shown when the Yiddish is romanized with the LC table for Yiddish; however, some LC romanization tables for languages written in Arabic script do make provision for the marking of abnormal positional forms of letters.

Additional Distinctions

Distinctions not present in the original written text may be introduced by romanization. Wellisch gives the example of the same Cyrillic script word that is rendered as *tsentral* (with the Library of Congress romanization table for Russian) and central (with the table for Serbian).[15] The same Hebrew word may be rendered differently when used in a Yiddish context. Many other examples can be cited: the same East Asian ideograph will be rendered differently by the Wade-Giles (Chinese), modified Hepburn (Japanese), or McCune-Reischauer (Korean) romanization systems sanctioned by the Library of Congress. For example, figure 2 shows three titles that begin with the same two ideographs, which are romanized as han tzu (Chinese), kanji (Japanese), and hancha (Korean). Only when the ideographs are read according to a particular pronunciation are the differences apparent.

Proliferation of Transcription Schemes

The difficulties for readers are compounded because various transliteration methodologies have been developed by various authorities for the same language and script. Hamdy examined five systems for the transliteration of the Arabic alphabet, and found that only thirteen of the twenty-eight letters were rendered the same way in all systems.[16]

The romanization tables of the Library of Congress have been approved by the American Library Association and the National Information Standards Organization (NISO) as standards for bibliographic transcription.[17] The tables are also used to create an *AACR2* heading when there is no well-known English form of a name; when a well-known English

```
Hs˘u, Ming, fl. 1973-
    (Han tzu ch'ang shih)
    漢字　常識　/徐　明　編. -- 香港　：進修　出版社，1973.
    40 p. : ill. -- (語文　小叢書)

    1. Chinese characters. 2. Chinese characters--History. I. Title. II. Series:
  Y˘u wen hsiao ts'ung shu (Hong Kong)


(Kanji naritachi jiten)
漢字　なりたち　辞典　：藤堂　方式・小学生版　/藤堂　明保　監修　；教育社　編.　--
    [東村山市]：教育社　：東京　：販売　教育社　出版　サービス，1982 (1985
    printing)
    1071 p. : ill. ; 22 cm.

    1. Chinese characters--Dictionaries--Japanese. 2. Chinese language--
  Etymology. 3. Japanese language--Study and teaching (Elementary)--Japan. I.
  T˘od˘o, Akiyasu, 1915- II. Ky˘oikusha.


Yu, Ch˘ong-gi, 1910-
    (Hancha chilli kyemong yoj˘on)
    漢字　眞理　啓蒙　要典　/柳　正基　著. -- 서울　：亞細亞　文化社，1988.
    428 p. : ill. ; 23 cm. -- (弘道　全書　; 11)

    1. Chinese characters--Korea--Dictionaries. I. Title. II. Title: Kyemong
  yoj˘on. III. Series: Hongdo ch˘ons˘o ; 11.
```

Figure 2. The Character for Character.

form does exist, it is quite possibly in a different romanization scheme!

Although the Library of Congress tables are widely used in the English-speaking world, they are not used universally. Librarians in other linguistic environments may prefer other romanization systems; for example, the transliteration standards developed by the International Organization for Standardization (ISO).

This Babel of competing transliterations, which impedes the sharing of cataloging data, occurs not only in the library world. The scholarly exchange of information is likewise impeded by the use of different transliteration systems, and the reference librarian must cope with all of them—transliteration systems used by libraries (the preservers of information) and different systems used by the originators of information.

Effect of Normalization

The last factor related to the defects of romanization is due to the organization of the catalog. In both the online catalog and its card catalog predecessor, diacritical distinctions (that have been deliberately included in the romanized text to show significant differences) are considered unimportant for retrieval, and so are usually ignored. The Technical Standards for Library Automation Committee (TESLA) survey on the indexing of OPACs included questions on two diacritical marks and two letter/diacritic combinations, but did not ask about diacritical marks or special characters important to romanization.[18] For example, there was no question on the romanized representations of the Hebrew and Arabic letters *ayn* and *alif* (the former is also used to transcribe the aspiration mark of Wade-Giles romanization). The survey found that diacritical marks are often dropped; the letter/diacritic combinations are either dropped outright or are converted to unaccented letters.

Most of LC's romanization schemes use both diacritical marks and special characters (for example, the romanized *ayn*). Normalization such as that documented in the TESLA survey has the following effect on retrieval via romanized access points: when two words differ only in diacritical marks and/or special characters, they cannot be distinguished. But this problem is not confined to romanized text: it also applies to accented languages written in Latin script. In their paper on bilingual (French/English) subject access, Rolland-Thomas and Mercure say, "Lack of accents in

```
066      ‡c$1
100 10   ‡6880-01‡aShen, Wei-pin.
880 10   ‡6100-01/$1‡a沈　渭滨.
245 10   ‡6880-02‡aHung Jen-kan /‡cShen Wei-pin chu.
880 10   ‡6245-02/$1‡a洪　仁玕 /‡c沈　渭滨　著.
250      ‡6880-03‡aTi 1 pan.
880      ‡6250-03/$1‡a第 1版.
260 0    ‡6880-04‡aShang-hai :‡bShang-hai jen min ch'u pan she :‡bHsin hua shu
     tien Shang-hai fa hsing so fa hsing,‡c1982.
880 0    ‡6260-04/$1‡a上海 :‡b上海人民出版社 :‡b新華書店上海发行所发行,‡c1982.
300      ‡a136 p., [1] leaf of plates :‡bill. ;‡c19 cm.
490 1    ‡6880-05‡aChung-kuo chin•tai shih ts'ung•shu
880 1    ‡6490-05/$1‡a中國　近代　史　叢書
504      ‡aIncludes bibliographical references.
600 10   ‡6880-06‡aHung, Jen-kan,‡d1822-1864.
880 14   ‡6600-06/$1‡a洪　仁玕,‡d1822-1864.
651  0   ‡aChina‡xHistory‡yTaiping Rebellion, 1850-1864.
650  0   ‡aRevolutionists‡zChina‡xBiography.
830  0   ‡6880-07‡aChung-kuo chin•tai shih ts'ung•shu (Shanghai, China)
880  0   ‡6830-07/$1‡a中國　近代　史　叢書 (Shanghai, China)
```

Figure 3. Record Containing Alternate Graphic Representation.

a keyword search could yield unwanted noise."[19] The fundamental rule of normalizing to ASCII ensures maximum retrieval, but can bring up unwanted matches. Perhaps OPACs should have an optional feature to eliminate unwanted matches of this sort by filtering the retrieval result to eliminate records that do not have the diacritical marks and special characters included in the search statement.

NONROMAN DATA IN USMARC

The factors described above show why original script access is far preferable for the reader of a foreign language, particularly when the language is written in a nonroman script. The first implementation of nonroman scripts in automated library systems was overseas: the Israeli ALEPH system with Hebrew and Latin scripts was introduced in 1981[20]; Cyrillic and Arabic have since been added to it.[21] The MINISIS library system has been Arabized, and completion of the adaptation of DOBIS/LIBIS for use in an Arabic environment was expected in 1990.[22]

The addition of nonroman scripts to USMARC had been considered as early as 1976,[23] but no modifications to USMARC were required until nonroman data became a reality with the implementation of East Asian scripts on RLIN in 1983. Subsequent American implementations of nonroman scripts have been East Asian scripts on OCLC in 1986,[24] and on RLIN, Cyrillic in 1986, Hebrew in 1988,[25] and Arabic in 1991.[26] Figure 3 shows a record that contains alternate graphic representation (Chinese characters).

General Principles

Despite the diversity of scripts that now can be included in USMARC bibliographic records, the underlying design is constant, and is used in the *USMARC Format for Authority Data* as well as in the bibliographic format.[27] The design has three significant parts:

1. Identification of the character sets present in the record;
2. Linked representations, in transliteration and in the original script; and
3. Allowable mixing of roman and nonroman scripts (one or more) in a single data element.

The identification of character sets and the mixing of scripts is based upon a standard methodology for character set extension, ISO 2022 (the comparable American national standard is ANSI X3.41).[28] Under this methodology, a

character set is identified and announced by an escape sequence (so called because it begins with the character escape, ASCII hex 1B). The mixing of scripts in the same record element is possible because each run of a particular script is unequivocally identified by the escape sequence that precedes it.

USMARC identifies each character set with a three-byte escape sequence. The initial character is escape; the intermediate character indicates whether the character set is single byte or multibyte; the final character identifies the character set itself. The European Computer Manufacturers' Association (ECMA) maintains a registry of character sets and assigns escape sequence codes for them. Privately defined sets may also be established by agreement between cooperating parties; ISO 2022 specifies escape sequences for such character sets. (The use of privately defined character sets in bibliographic records will be discussed below.)

Latin script is the default script for USMARC; the ASCII and ANSEL character sets are not identified in the record.[30,31] (The upper range of the eight-bit USMARC Latin character set is actually a subset of the standard ANSEL character set; *Proposal 92–6*, which was discussed by MARBI at its January 1992 meeting, addresses addition of the missing characters.[32]) Only nonroman scripts that are present in the record are explicitly identified, within the 066 field, "Character sets present." Each nonroman character set is identified in an occurrence of subfield c; the subfield contains the intermediate and final character of the escape sequence for the character set. The presence of the 066 field shows that there is nonroman data in the USMARC record. The character sets that have been used to transcribe the nonroman data in the record are identified in the occurrences of subfield c in the 066 field.

The nonroman data themselves are carried in multiple occurrences of the 880 field, "Alternate graphic representation." The function of each 880 field is shown by a linking tag. In most cases, a nonroman field (or alternate graphic representation) matches a romanized field in the record, but USMARC does support alternate graphic representation without requiring a corresponding romanization.

The information that explicitly links two fields is carried in subfield 6, defined for the 880 field and for any field in the range 0XX–8XX that may have an alternate graphic representation. The linkage information in subfield 6 belonging to one field consists of the tag of the other field and an arbitrarily assigned occurrence number (to distinguish this field pair from other pairs with the same tag combination). When the record does not contain a romanized field to which an alternate graphic representation can be linked, subfield 6 of the 880 field contains the tag of a hypothetical romanized field and the occurrence number 00. For example, if a library chooses not to include the romanized equivalent of a note itemizing the contents of a Chinese-language publication, the 880 field holding the transcription of the contents information would be "linked" to a nonexistent field, the romanized contents note that could have been made, but was not.

Currently Supported Nonroman Scripts

Nonroman character sets specified for use in USMARC cover Hebraic, Arabic, and Cyrillic letters; Chinese characters; Japanese kana; and Korean hangul.[33] This repertoire includes all the characters required for original script transcription of JACKPHY language material.

USMARC specifies the use of a number of standard character sets: ASCII, ANSEL, the American standard East Asian Character Code (EACC),[34] a Gosudarstvennyi Komitet SSSR Po Standartam (GOST) standard from the former Soviet Union for the Russian alphabet, and the international standard for an extended Cyrillic character set for other Slavic languages.[35] The privately defined basic Arabic character set is an augmented version of an existing standard, ISO 9036,[36] which is taken from Arab Standard ASMO 449.[37]

EACC, although an American national standard, is still defined as a privately defined character set in USMARC records. Hebrew, basic Arabic, and extended Arabic are also privately defined sets. The source of this private definition lies in the use of RLIN by the Library of Congress for nonroman cataloging. The Research Libraries Group (RLG) implemented these scripts as privately defined character sets specific to RLIN. Registering the character sets with ECMA might have delayed implementation of the various scripts.

The Library of Congress adopted the RLIN escape sequences for USMARC, but

has said it will apply for ECMA codes for ANSEL and EACC. The USMARC Hebrew character set and the extended Arabic characters used in USMARC records have been proposed as ISO standards (so would be registered with ECMA upon approval as international standards). All the USMARC character sets should probably have unequivocal ECMA registration, since there is no guarantee that a privately defined escape code will have the same meaning in all contexts.

The USMARC format is based on ISO 2022 rules, which means that (in theory) any ECMA-registered character set could be exchanged in USMARC records, without its formal promulgation as a USMARC character set by the Library of Congress. The USMARC Specifications includes as an example[38] the use of the escape sequence for the ISO Greek bibliographic standard.[39] Privately defined character sets are also permitted under ISO 2022 rules.

The addition of nonroman scripts to local systems is occurring. Innovative Interfaces released East Asian script support for the INNOPAC system in June 1991.[40] Geac and VTLS have contracts to develop systems with Cyrillic capability: Geac for the Academy of Sciences in Leningrad, and VTLS for the Lenin Library in Moscow.[41,42] Dynix has a contract with the Santa Clara County Library in California that includes a requirement for East Asian scripts.[43]

As long as there continues to be close cooperation between the libraries and agencies operating in the area of nonroman scripts, privately defined character sets (and their escape codes) do not present a significant problem in data exchange; the use of character set standards that have not been authorized for USMARC use is unlikely. The registration of all USMARC character sets is desirable, because it will establish standard reference codes for character sets in all multiscript bibliographic applications, including the exchange of USMARC bibliographic data internationally.

A GLOBAL CHARACTER SET

The problems of a plethora of character sets and their unequivocal identification are not unique to libraries. All multiscript applications have this problem. It is also inefficient to build software for discrete language or script contexts. Recognition of the problems of multiple character sets shows the need for a single global character set. Over the last few years, there have been two efforts directed towards the definition of a comprehensive character set to encode the principal writing systems of the world: Unicode, a project based in California, and 10646, a standards activity of Joint Technical Committee 1 of the ISO and the International Electrotechnical Commission.[44]

RLG has participated in both developments. Since 1989, RLG has been a member of the ANSI X3L2 subcommittee, the United States Technical Advisory Group (TAG) to the Joint Technical Committee 1/Subcommittee 2 (JTC1/SC2), which is responsible for the development of 10646. Each national standards body has a group designated as the national focal point for review of the various stages of 10646: the U.S. focus is the ANSI subcommittee X3L2, and the Canadian focus is Standards Council of Canada, Subcommittee 2.

RLG is a founding member of the Unicode Consortium, the nonprofit organization established (under the name Unicode, Inc.) to develop and promote Unicode, a sixteen-bit character-encoding standard.[45] Since 1989, the author, with two colleagues at the RLG, has contributed a bibliographic perspective to the deliberations of the Unicode Technical Committee and its informal predecessor, the Unicode Working Group.

Unicode

Work on Unicode began in 1988, when experts at Apple Computer Corp. and Xerox Inc. envisioned a multiscript character set with the simplicity and power of ASCII, an ASCII "stretched to 16 bits to encompass the characters of all the world's living languages."[46]

A fixed width of sixteen bits provides for a universe of more than 65,000 unique codes and extends the benefits of ASCII to a multiscript environment. The Unicode standard's other fundamental principle is that it encodes characters, not higher-level (and far more numerous) "textual elements," presentation forms, glyphs, and typographical distinctions. (These higher-level visualizations of characters may be encoded in other standards, for example, the AFII glyph registry.) Figure 4 shows the distribution of characters within the Unicode codespace; each square on this grid contains 256 code points.

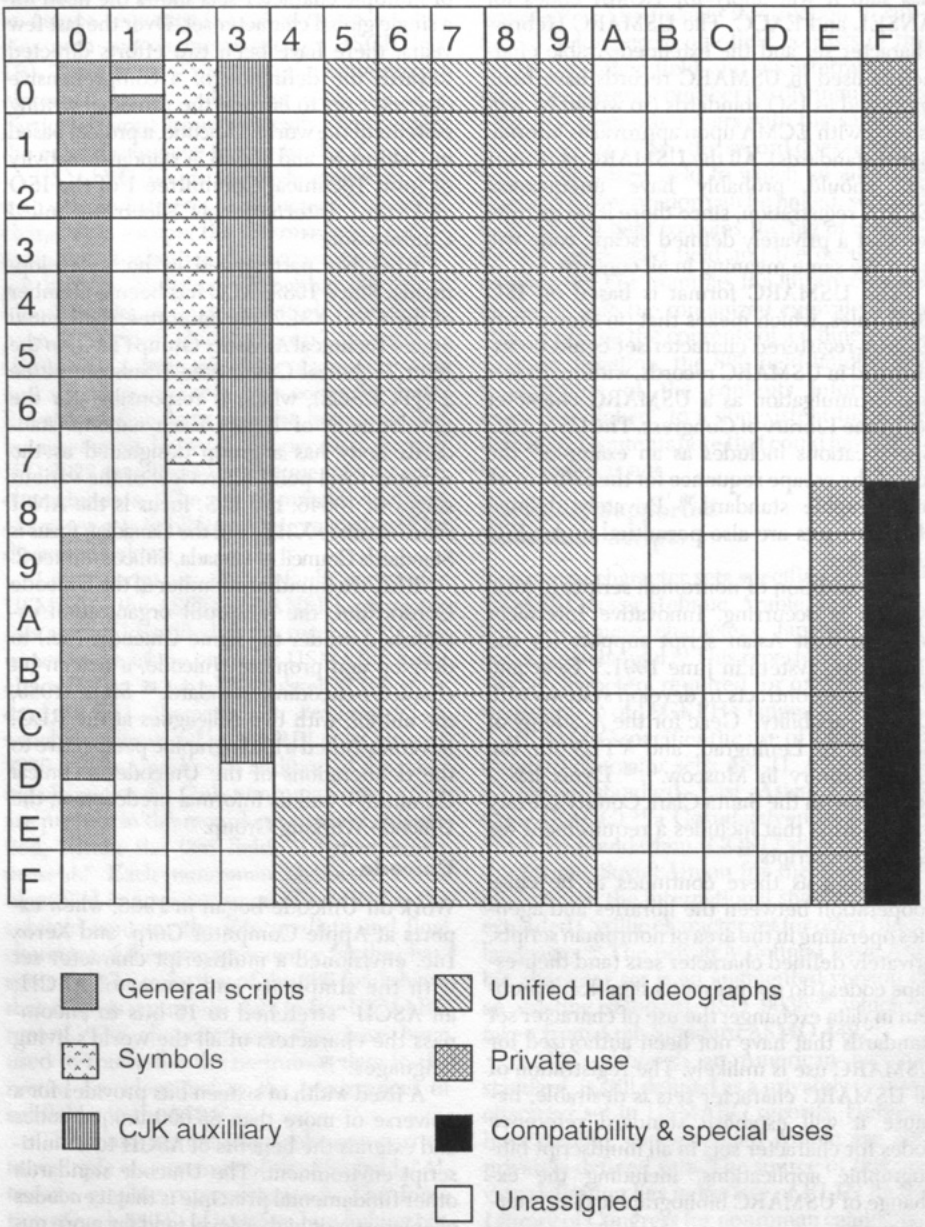


Figure 4. Distribution of Characters within Unicode Codespace.

The Unicode standard incorporates many coding concepts familiar to librarians, including nonspacing characters (e.g., diacritical marks) and Han unification. Some combinations of letters and diacritical marks are also encoded, for compatibility with other character set standards. (Readers may be familiar with Microsoft's "ANSI standard" character set for Windows,[47] which is based upon such a standard, International Standard ISO 8859-1, Latin1.[48])

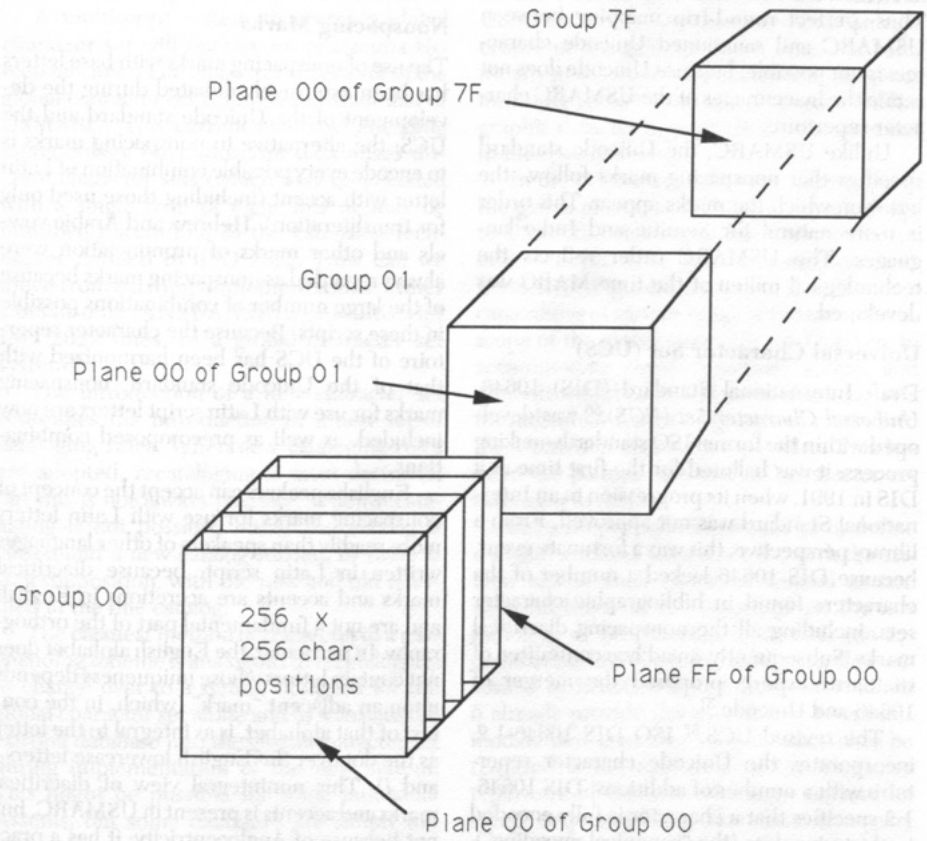


Figure 5. After Figure 1 in ISO/IEC DIS 10646-1.2.

"Han unification" is the term used to describe a union of the ideographic characters used in writing Chinese, Japanese, and Korean. A single ideograph often has the same meaning in all three languages, although the verbalization associated with it is quite different. Han unification can be illustrated by the two ideographs, the name of the Chinese Han dynasty and the ideograph for "word," that (in combination) represent the concept "characters." The two ideographs begin each of the titles in figure 2. The romanized title, enclosed in parentheses, shows how the identical character pair is read in different languages: han tzu in (Mandarin) Chinese, kanji in Japanese, and hancha in Korean.

Han unification has been an element of library data processing since 1983, when RLIN CJK was released. EACC (formerly REACC) is a composite character set,[49] encoding Japanese kana, Korean hangul, and the ideographs used to write Chinese, Japanese, and Korean. In EACC, each ideograph is encoded only once, not once for Chinese, again for Japanese, and a third time for Korean. The millions of East Asian records in RLIN and OCLC are proof that Han unification works. In 1990, IBM sponsored an examination of the issues of Han unification at the University of Toronto, under the direction of Professor Kazuko Nakajima. This study (as yet unpublished) independently confirmed the appropriateness of Han unification as an approach to the encoding of East Asian scripts.

The Unicode standard incorporates all the characters of the USMARC character sets, except where these character sets are in error; for example, a number of Chinese characters,

which were incorrectly assigned unique codes in EACC, are encoded only once in Unicode. Thus, perfect round-trip mapping between USMARC and sanctioned Unicode characters is not possible, because Unicode does not retain the inaccuracies of the USMARC character repertoire.

Unlike USMARC, the Unicode standard specifies that nonspacing marks follow the letter on which the marks appear. This order is more natural for Semitic and Indic languages. The USMARC order reflects the technological milieu at the time MARC was developed.

Universal Character Set (UCS)

Draft International Standard (DIS) 10646, *Universal Character Set (UCS)*,[50] was developed within the formal ISO standards-making process; it was balloted for the first time as a DIS in 1991, when its progression to an International Standard was not approved. From a library perspective, this was a fortunate event, because DIS 10646 lacked a number of the characters found in bibliographic character sets, including all the nonspacing diacritical marks. Subsequently, an ad hoc committee of standards experts proposed the merger of 10646 and Unicode.[51]

The revised UCS,[52] ISO DIS 10646-1.2, incorporates the Unicode character repertoire with a number of additions. DIS 10646-1.2 specifies that a character is fully encoded in thirty-two bits (the "canonical encoding"), but a sixteen-bit abbreviated form is permitted for characters from the Basic Multilingual Plane. The organization of the 10646 codespace is shown in figure 5. In the current version of 10646, only character encodings for the Basic Multilingual Plane are proposed.

Draft International Standard 10646 did not include a repertoire of East Asian ideographs; much of DIS 10646-1.2 consists of tables assigning codes to specific ideographs. JTC1/SC2 appointed a Chinese/Japanese/Korean Joint Research Group (CJK JRG), to study the encoding of ideographs and, in particular, the issue of unification. The principal countries and regions in Asia were represented, as was the United States. The ideographic tables in DIS 10646-1.2 are the direct result of agreements reached by the CJK JRG; they are based on the "unified Han" character repertoire developed by standards experts from the People's Republic of China and members of the Unicode Technical Committee.

Nonspacing Marks

The use of nonspacing marks with base letters has been extensively debated during the development of the Unicode standard and the UCS; the alternative to nonspacing marks is to encode every possible combination of Latin letter with accent (including those used only for transliteration). Hebrew and Arabic vowels and other marks of pronunciation were always regarded as nonspacing marks because of the large number of combinations possible in these scripts. Because the character repertoire of the UCS has been harmonized with that of the Unicode standard, nonspacing marks for use with Latin script letters are now included, as well as precomposed combinations.

English speakers can accept the concept of nonspacing marks for use with Latin letters more readily than speakers of other languages written in Latin script, because diacritical marks and accents are accretions in English and are not a fundamental part of the orthography. In particular, the English alphabet does not contain letters whose uniqueness depends upon an adjacent "mark" (which, in the context of that alphabet, is as integral to the letter as the dot over the English lowercase letters *i* and *j*). This nonintegral view of diacritical marks and accents is present in USMARC, but not because of Anglocentricity; it has a practical origin. A limited repertoire of nonspacing marks together with the ASCII letters allows a much larger repertoire of "marked" Latin letters to be encoded with a single extended ASCII (eight-bit) character set.

Because of the influence of American-developed MARC, libraries everywhere have accepted the use of nonspacing Latin script marks; this acceptance is exemplified in international bibliographic character set ISO 5426 and the many national MARC character sets, all of which are modeled on USMARC.[53,54]

Anticipating the Future

The outcome of the ISO DIS 10646-1.2 ballot is unknown. A single global character set is certainly preferable to two. Regardless of the outcome, there will eventually be systems with a greatly extended character repertoire (and associated fonts for presentation of the data). We must plan for the use of such

systems in libraries, and consider how a global character set might interact with USMARC.

A multiscript system that uses a global character set will operate on characters sixteen bits long. The character repertoire of the global character set is far greater than that of USMARC. The current standard encoding methodology for multiscript data mixes discrete character sets (which may be encoded with a single byte of eight bits or may be multibyte) according to the extension techniques defined by ISO 2022. How can we move from existing implementations, either Roman-only (ANSI/ANSEL) or multiscript ISO 2022 ones, to a global character set system?

The introduction of a new character set resembles the introduction of a new set of cataloging rules. When new cataloging rules are adopted, recataloging is most desirable, but this is not always possible. When recataloging is not possible, either the catalog is closed and a new catalog started, or there is superimposition, with the new and old interfiled in the one catalog.

The cleanest method of transition to a new encoding scheme is analogous to recataloging: to change over to a system designed for the global character set alone and to translate the existing database into the global character set during implementation of the new system. This might be possible for some individual systems, but some databases are simply too massive to convert. In this situation, the only solution is to superimpose the new character set on the old, mapping between the two sets. (Such mapping is analogous to that done when ASCII peripherals are used with an EBCDIC mainframe; the Extended Binary Decimal Interchange Code [EBCDIC] was developed by IBM and is used in its mainframe computers.)

Mapping to and from the Global Character Set

Characters from the current USMARC character repertoire can be mapped to sanctioned global characters, except in a few cases. For example, a few EACC characters are now known to be duplicative, and so have not been included in the new "unified Han" standard (part of DIS 10646-1.2, and vol. 2 of *The Unicode Standard*). These EACC characters would have to be mapped to values in the private-use zone if they were to be transmitted unchanged.

There will need to be a specification to dictate mapping between the current USMARC character repertoire and the repertoire of the global character set, and rules for any problematic characters. Mapping may introduce some irreversible changes to bibliographic data; these must also be documented in the specification.

In the other direction, some characters in the global character set have exact mappings to USMARC characters, but there are many other scripts that are not supported in USMARC. Figure 6 compares the character capabilities of various types of system with the scope of the global character set. One way to accommodate these "residual" scripts and symbols would be to add many character sets for individual scripts or groups of symbols to the USMARC character repertoire. This has been the pattern of USMARC character set extension to date.

But a simpler solution would be to define a single "residual" character set, with the sixteen bits of each character treated as two eight-bit bytes (or two octets, to use the terminology of DIS 10646). USMARC can accommodate multibyte encodings, as demonstrated by EACC. The 066 field and subfield 6 already provide the mechanism to accommodate new character sets; all that would be required is the definition of the character repertoire of this new character set and an escape sequence for it. The character repertoire could be defined as the global character set minus those characters having a specified USMARC character set mapping.

Data exchange between systems of different script capability already occurs. For example, when a Roman-only system receives data from a multiscript system, the unusable nonroman data are either stored (in anticipation of future use) or discarded. The presence of nonroman data is shown by the 066 field, "Character sets present." In the same way, data from new "universal" systems, which use the global character set and have additional scripts not found in USMARC, can be partially utilized by today's systems. (The term *universal system* should not be confused with Spalding's single-alphabet "universal catalog" mentioned at the beginning of this paper.)

The objective of global character set development is to eliminate the need for many character sets in multiscript processing. Using

ISO 2022 techniques with a global character set violates the whole premise of universality and forces USMARC to cling to an old standard that will become increasingly obsolete in the world of the global character set. Treating the sixteen-bit codes as "double bytes" may cause processing difficulties; for example, if a particular eight-bit pattern within a sixteen-bit character code corresponds to a control character in the eight-bit world. But current systems and data cannot be abandoned; initially, the global character set must be accommodated within the existing environment.

Moving Away from ISO 2022

A later phase in the integration of the global character set into USMARC might be to make the global character set the preferred character set for alternate graphic representation. Library needs were not ignored by the compilers of the Unicode standard; library vendors will soon be using Unicode-based products in library systems. The development of the global character set is a radical break with the past, so USMARC must change, sooner or later. Preferring the global character set to multiple character sets is the first step towards moving USMARC away from the ISO 2022 model.

Individual systems may continue to use their own internal representation of data, but they will have to be able to map their nonroman character repertoire to and from the global character set. The onus would be placed on the obsolescent system to examine incoming nonroman data and determine whether they contained any unsupported characters.

Use of the global character set in USMARC is not unreasonable. The 066 field, "Character sets present," would have to be modified to show that the sixteen-bit global character set was being used for alternate graphic representation (instead of the ISO 2022 technique). Perhaps one of the indicators of the 066 field could be defined to encode two values: use of the single global character set or use of the ISO 2022 model. In the latter case, subfield c would carry the initial and final characters of the escape sequence of every character set in the record. For processing purposes, the sixteen bits of the global character set might still be handled as two eight-bit bytes, since the default character set would still be eight-bit ASCII/ANSEL.

As systems that can utilize the global character set directly proliferate, the Library of Congress might offer USMARC records completely converted to sixteen-bit encoding, as an option to the conventional eight-bit, ISO 2022 encoding. At this point, the USMARC formats will have to include an identifier for the type of encoding in the record as a whole. Subfield a of the 066 field defines the default G0 character set, but this definition is based on the ISO 2022 model. A type of encoding

Script	Roman Only	East Asian	Semitic/ Cyrillic	All USMARC	Global
Latin	Latin	Latin	Latin	Latin	Latin
East Asian	NO	East Asian	NO	East Asian	East Asian
Cyrillic	NO	NO	Cyrillic	Cyrillic	Cyrillic
Hebrew	NO	NO	Hebrew	Hebrew	Hebrew
Arabic	NO	NO	Arabic	Arabic	Arabic
Devanagari	NO	NO	NO	NO	Devanagari
Thai	NO	NO	NO	NO	Thai
Armenian	NO	NO	NO	NO	Armenian
etc.	NO	NO	NO	NO	*etc.*

Examples of System Types

Roman Only: WLN, NOTIS, many others
East Asian: OCLC, INNOPAC
Semitic/Cyrillic: ALEPH
All USMARC: RLIN

Figure 6. The Global Character Set and Current Capabilites.

identifier is fundamental information for processing, so would be more appropriate as an announcer at the beginning of a record (or possibly a file).

The final phase in the use of the global character set in the library environment will be the abandoning of eight-bit USMARC as an exchange medium and the use of the global character set for *all* the data in a machine-readable bibliographic record, not just for the alternate graphic representation. This step cannot come about until universal systems are common in libraries, early in the next millennium if not sooner. Until this time, nonroman USMARC records will contain eight-bit ASCII/ANSEL and other character encodings.

At the present time, catalogers romanize the bibliographic description so that records for nonroman script material can be seen on devices limited to Latin script. Romanization is a time-consuming process and frequently requires detailed knowledge of the language written in the nonroman script. Once multiscript universal systems are widely available, the need for a romanized bibliographic description will disappear. The alternate graphic representation will become the only bibliographic description, and, at long last, cataloging rules and machine-readable cataloging will be in harmony.

CONCLUSION

The future of character set standards is far from clear. At the time of writing, the revision of DIS 10646 that incorporates the Unicode character repertoire and values as the Basic Multilingual Plane is being balloted. Whether there will be two global character set standards or only one remains to be seen. Regardless of the outcome, however, systems librarians need to plan for the day when nonroman scripts will be as readily available as ASCII is today.

REFERENCES

1. *Anglo-American Cataloguing Rules*, 2d ed., prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association, the Australian Committee on Cataloguing, the British Library, the Canadian Committee on Cataloguing, the Library Association, the Library of Congress; ed. Michael Gorman and Paul W. Winkler (Ottawa: Canadian Library Assn.; Chicago: American Library Assn., 1988).
2. Network Development and MARC Standards Office, *USMARC Format for Bibliographic Data*, including guidelines for content designation (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1988–).
3. Karen T. Wei, "Current Status and Future Trends of East Asian Library Automation in North America." *Information Technology and Libraries* 5, no.2:140–46 (June 1986).

   RLIN and CJK are trademarks and service marks of the Research Libraries Group, Inc.
4. C. Sumner Spalding, "Romanization Reexamined," *Library Resources & Technical Services* 21, no.1:3–12 (Winter 1977).

   Correspondence, *Library Resources & Technical Services* 21, no.3:303–5 (Summer 1977). The quotation will be found on p.8 of Spalding's paper.
5. Bella Weinberg, "Transliteration in Documentation," *Journal of Documentation* 30, no.1:18–31 (Mar. 1974).
6. Spalding, "Romanization Reexamined."
7. Hans W. Wellisch, "Script Conversion and Bibliographic Control of Documents in Dissimilar Scripts: Problems and Alternatives," *International Library Review* 10:3–22 (1978).
8. S. Michael Malinconico and Walter R. Grutchfield; Erik J. Steiner, "Vernacular Scripts in the NYPL Automated Bibliographic Control System," *Journal of Library Automation* 10, no.3:205–25 (Sept. 1977).
9. Ibid., p.209.
10. Alan W. Watts, *The Way of Zen* (New York: Vintage, 1989), p.xvi.
11. John Eilts, personal communication.
12. Laura Spurrier, personal communication.
13. *The Pinyin Chinese-English Dictionary* (Beijing: Commercial Pr., 1979), p.710–11.
14. Bella Hass Weinberg, "Hebraica Cataloging and Classification" in *Cataloging and Classification of non-Western Material: Concerns, Issues and Practices*, ed. Mohammed M. Aman (Phoenix, Ariz.: Oryx, 1980), p.321–57.
15. Wellisch, "Script Conversion," p.12.
16. Nabil Hamdy, "Cataloging and Classifying Arabic Materials," in *Cataloging and Classification of non-Western Material: Concerns, Issues and Practices*, ed. Mohammed M. Aman (Phoenix, Ariz.: Oryx, 1980), p.268–96.
17. *ALA-LC Romanization Tables: Transliteration Schemes for non-Roman Scripts*, approved by the Library of Congress and the American Library Association; tables comp.

and ed. Randall K. Barry (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1991–).

18. "Variations in Indexing Practices among OPACs," *Information Technology and Libraries* 10, no.1:67–73 (Mar. 1991).
19. Paule Rolland-Thomas and Gerard Mercure, "Subject Access in a Bilingual Online Catalogue," *Cataloging & Classification Quarterly* 10, nos.1/2:141–49 (1989).
20. Judith Levi, "ALEPH—An Online Real-Time Integrated Library System," *Judaica Librarianship* 1, no.2:58–63 (Spring 1984).
21. Susan S. Lazinger, "ALEPH: Israel's Research Library Network: Background, Evolution, and Implications for Networking in a Small Country," *Information Technology and Libraries* 10, no.4:275–91 (Dec. 1991).
22. Abdus Sattar Chaudhry and Mohammad Saleh Ashood, "Potential of DOBIS/LIBIS and MINISIS for Automating Library Functions: A Comparative Study," *Program* 24, no.2:109–28 (Apr. 1990).
23. James Agenbroad, "Inclusion of Nonroman Characters in the MARC and UNIMARC Formats," Report to MARBI, 1976.
24. Karen T. Wei and Sachie Noguchi, "RLIN CJK versus OCLC CJK: The Illinois Experience," *Library Resources & Technical Services* 33, no.2:140–51 (Apr. 1989).
25. Joan M. Aliprand, "Hebrew on RLIN," *Judaica Librarianship* 3, nos.1/2:5–16 (1986–87).
26. "LC Is First to Use New RLIN Arabic-Script Cataloging System," Library of Congress Information Bulletin 50, no.22:445 (Nov. 18, 1991).
27. Network Development and MARC Standards Office, USMARC Format for Authority Data, Including Guidelines for Content Designation (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1987–).
28. International Organization for Standardization, Information Processing—ISO 7-Bit and 8-Bit Coded Character Sets—Code Extension Techniques, 3d ed. (ISO 2022:1986) (Geneva: International Organization for Standardization, 1986).
29. American National Standards Institute, *Code Extension Techniques for Use with the 7-Bit Coded Character Set of American National Standard Code for Information Interchange* (ANSI X3.41-1974) (New York: ANSI, 1974).
30. American National Standards Institute, *Coded Character Set—7-Bit American National Standard Code for Information Interchange* (ANSI X3.4-1986) (New York: ANSI, 1986).
31. American National Standards Institute, *Extended Latin Alphabet Coded Character Set for Bibliographic Use* (ANSI Z39.47-1985) (New York: ANSI, 1985).
32. "Additions to the USMARC Character Set," *Proposal No. 92–6* (Washington, D.C.: Network Development and MARC Standards Office, Library of Congress, 1991).
33. Network Development and MARC Standards Office, *USMARC Specifications for Character Sets, Record Structure, Tapes* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1990–).
34. American National Standards Institute, *East Asian Character Code for Bibliographic Use* (ANSI Z39.64-1989) (New Brunswick: Transaction, 1990).
35. International Organization for Standardization, *Extension of the Cyrillic Alphabet Coded Character Set for Bibliographic Information Interchange* (ISO 5427:1984) (Geneva: ISO, 1984).
36. International Organization for Standardization, *Information Processing—Arabic 7-Bit Coded Character Set for Information Interchange* (ISO 9036:1987) (Geneva: ISO, 1987).
37. Arab Organization for Standardization and Metrology, *Data Processing—7-Bit Coded Arabic Character Set for Information Interchange* (Arab Standard Specifications, no.449-1982 authorized English translation issued by ASMO), [s.l.] AMSO, 1983.
38. Network Development and MARC Standards Office, USMARC Specifications, p.18.
39. International Organization for Standardization, *Greek Alphabet Coded Character Set for Bibliographic Information Interchange* (ISO 5428:1984) (Geneva: ISO, 1984).
40. Innovative Interfaces, Inc., "Innovative Interfaces Offers CJK Support," press release dated June 18, 1991.
41. Ian D. Manson, personal communication.
42. "VTLS Signs State Lenin Library," *Library Systems Newsletter* 9, no.6:46 (June 1989).
43. Kim Schmidt, personal communication.
44. Kenneth M. Sheldon, "ASCII Goes Global," *BYTE* 16, no.7:108–16 (July 1991).
45. Unicode Consortium, *The Unicode Standard: Worldwide Character Encoding*, ver. 1.0 (Reading, Mass.: Addison-Wesley, 1991–).
46. Joseph D. Becker, "Unicode 88," unpublished paper dated August 29, 1988.
47. *CompuAdd Windows: User's Guide for the Windows Environment, Version 3.0 for MS-DOS Operating System*. Rev. ed. (Austin, Tex.: CompuAdd Corporation, 1990), p.567–68.
48. International Organization for Standardization, Information Processing-8-Bit Single-Byte Coded Graphic Character Sets: Part 1: Latin alphabet no. 1 (ISO 8859-1:1987) (Geneva: ISO, 1987).
49. Karen Smith-Yoshimura and Alan Tucker, "RLIN East Asian Character Code and the CJK Thesaurus," in *Proceedings of the Second*

Asian-Pacific Conference on Library Science: 20-24 May, 1985, Seoul (Seoul, Korea: Cultural and Social Centre for the Asian and Pacific Region, 1985), p.105–37.

50. International Organization for Standardization/International Electrotechnical Committee, *Information Technology—Universal Coded Character Set* (UCS) ([s.l.] ISO/IEC Draft International Standard 10646, 1990).
51. Edwin Hart, "10646M (Merger) Discussion Final Minutes: Summary of Results of Informal Meeting to Discuss Merging of DIS 10646 and Unicode into One Code," document 10646M/91-01, dated May 30, 1991, e-mail distribution from HART@APLVM.Bitnet (or HART@APLVM.JHUAPL.EDU).
52. International Organization for Standardization/International Electrotechnical Committee, *Information Technology—Universal Multiple-Octet Coded Character Set* (UCS): Part 1: Architecture and Basic Multilingual Plane, working document for ISO/IEC Draft International Standard 10646-1.2 ([s.l.] ISO/IEC DIS 10646-1.2, 1991).
53. International Organization for Standardization, *Extension of the Latin Alphabet Coded Character Set for Bibliographic Information Interchange*, 2d ed. (ISO 5426:1983) (Geneva: ISO, 1983).
54. John Clews, *Language Automation Worldwide* (Harrogate, Yorkshire: Sesame Computer Projects, 1988). ■■

The Development and Implementation of the USMARC Format for Classification Data

Rebecca S. Guenther

This paper discusses the newly developed USMARC Format for Classification Data. It reviews its potential uses within an online system and its development as one of the USMARC standards. It provides a summary of the fields in the format and considers the prospects for its implementation. The paper describes an experiment currently being conducted at the Library of Congress to create USMARC classification records and use a classification database in classifying materials in the social sciences.

The Library of Congress recently completed the development of a machine-readable format for classification data to allow for the communication of classification records between systems and to provide a standard for the storage of classification data in the computer. The *USMARC Format for Classification Data* joins the family of machine-readable cataloging (MARC) formats: bibliographic, authority, and holdings formats. Implementation poses great challenges for institutions, particularly for those responsible for the maintenance of library classification schemes.

POTENTIAL USES FOR ONLINE CLASSIFICATION

Online classification data have many potential uses for information access. They may provide the authority for classification numbers, terms, and shelflist information; they may be used for printing and maintaining a classification scheme; and they may enhance subject retrieval, assist the classifier, facilitate maintenance tasks for classification numbers in bibliographic records, and provide the basis for an online shelflist.

Authority Control for Classification Data

Online classification data may provide authority control for the classification number and caption(a heading that corresponds to a classification number(s) and describes the subject covered). An authoritative file of classification records may be used by the classifier to assign classification numbers to bibliographic records. It may also provide a system with the mechanism to validate the correct assignment of classification numbers.

In addition, it can provide authority control for synthesized classification numbers, i.e., numbers that have been made more specific by adding other numbers from a table or other parts of the schedule to a base number. A synthesized classification number need not appear in the classification scheme itself, since it is built by following add instructions, which instruct the classifier to add or append other numbers from the schedule or a table to a base number. Creating a classification record for a synthesized number can provide an authority for that number and facilitate its further use.

Rebecca S. Guenther is Senior MARC Standards Specialist, Network Development and MARC Standards Office, Library of Congress, Washington, D.C. This is an edited version of a paper presented at the fifth International Study Conference on Classification Research in Toronto, Canada, on June 25, 1991.

Printing and Maintenance of Classification Schedules

Online classification data could be an efficient method for printing a classification schedule. However, a print program for publishing the schedules will have different system requirements than the program for online display. Specifications will need to be developed when implementing an online classification system and print program.

The two major classification schemes in use in the United States, the Library of Congress Classification (LCC) and the Dewey Decimal Classification (DDC), have been developed, produced, and maintained very differently over the years. LCC is an enumerative scheme, with new classification numbers inserted where appropriate, and individual changes communicated through the publication *LC Classification—Additions and Changes*. DDC is hierarchical and uses number building extensively by appending numbers from other parts of the schedule onto a base number to create a more specific classification number. Revised editions of the whole scheme or of special sections have communicated changes to users; it is currently in its twentieth edition.

The *LCC*, now consisting of forty-six separate schedules, was developed over a period of time by different people. It was designed as a shelf location and browsing device and has been maintained as such.[1] Some of the schedules have been revised by using word processing software, although many have not. Most of the schedules, including many that are machine-readable because they are in WordPerfect, will require editing to make the references and notes more consistent with one another and to allow for the classification number records to be input into the USMARC format. Producing the *LCC* from an online file could facilitate the cumbersome process that the Library of Congress currently uses to publish revised editions. In addition, it may be used to make additions and changes to the classification schedules on a timely basis.

The *DCC* uses an online system to support the editorial process used to publish its classification scheme.[2] This system, called the Editorial Support System (ESS), does not provide the desired online access to classification numbers and terms that the USMARC format would. Because it is adequate to satisfy DDC's publication needs, it is not intended that the *USMARC Format for Classification Data* be used for printing the volumes of the *DDC*.

Providing Subject Access

In previous research, scholars and researchers recognized the possibilities of classification data providing subject access to library material. The DDC Online Project demonstrated the usefulness of classification data for subject access, browsing, and display, opening up potentially powerful new search strategies.[3] Explorations into the use of the LCC for online subject access have also been conducted.[4] Online classification data can enhance retrieval of bibliographic records by providing a different type of subject access, through a classed catalog approach, rather than through controlled subject headings. It enables the library user to see the interrelationships between topics and classification numbers and facilitates browsing from more general to more specific topics and numbers. In addition, more precise searching of bibliographic records is possible through classification numbers for certain types of searches. For example, a well-known individual author or work may have its own classification number in the *LCC*, and a search by class number may retrieve a listing of bibliographic records on that author or work more efficiently than a traditional author or title search. Having the classification data accessible online will facilitate this process.

Assistance for the Classifier

Online access to classification numbers may save time for the classifier. The classifier may perform a keyword search to locate quickly the possible classification numbers through captions, notes, or index terms. The classifier may be able to trace the formation of synthesized numbers to assist him in classifying. In addition, online classification may assist the classifier in obtaining consistency in the assignment of classification numbers to library materials. In the future, libraries may be able to develop expert systems that could do much of the work of applying add instructions or tables to classification numbers and validate the accuracy of synthesized classification numbers.

Maintenance of Classification Numbers in Bibliographic Records

An online system may use an automated classification database to facilitate a library's conversion to a different classification scheme. By maintaining the classification schedule online, an institution may be able to perform global updates of the bibliographic file when a classification number is added or changed.

Providing a Basis for an Online Shelflist

The Library of Congress has had numerous requests to automate its shelflist. If this project is undertaken, the online classification schedules would certainly provide a basis for that type of effort, or for another institution using LCC. An automated shelflist would be a valuable resource for bibliographic material.

USMARC AND THE ESTABLISHMENT OF BIBLIOGRAPHIC STANDARDS

The USMARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form. They are communication formats, primarily designed to provide specifications for the exchange of bibliographic and related information for sharing between systems on magnetic tape within the United States. Developers of the USMARC formats have attempted to maintain some compatibility with other national and international formats (e.g., CANMARC,UNIMARC). The Network Development and MARC Standards Office of the Library of Congress is responsible for developing and maintaining the USMARC formats, which now consist of the following: *USMARC Format for Bibliographic Data, USMARC Format for Authority Data,* and the *USMARC Format for Holdings Data.* The provisionally approved *USMARC Format for Community Information* will be published in 1992. Development and maintenance is accomplished in consultation with user communities.

Content designation in the USMARC formats are codes and conventions used to identify explicitly data elements in a record. The goal is to characterize the data elements with sufficient precision to support manipulation of the data for various functions.[5] Functions supported include display, both formatting in an online display and in producing a printed or other type of product, and online information retrieval. How an institution displays the data is not specifically covered in the USMARC formats, although they do provide for display constants, which are terms, punctuation, or spacing that are system generated for display. For instance, the hyphen separating the beginning and ending numbers of a classification number span is generated as a display constant by the system based on the structure ofthe classification number field of the record.

MARC (the generic term machine-readable cataloging; USMARC is the catalog used in the United States) has proven to be flexible, efficient, and easy to maintain for library automation systems.[6] It is used worldwide for storing, sharing, and manipulating bibliographic information by computer. The *USMARC Format for Classification Data* joins the other established formats with a similar structure and goals.

USMARC CLASSIFICATION FORMAT DEVELOPMENT

The Library of Congress recognized the need for a USMARC format for communicating and storing classification data and began the development of the USMARC classification format in 1987–88. It was developed in close consultation with the two major classification schemes in use in the United States, the LCC and the DDC, although its data elements are intended to be generic enough to accommodate other classification schemes. The Network Development and MARC Standards Office began work on a proposal to include classification data as an extension of the *USMARC Format for Authority Data.* The Library of Congress followed several assumptions in developing the format: certain types of data that need to be displayed or searched must be identified separately, all types of notes and classification numbers need to be accommodated by the format, and the schedules themselves should drive the format rather than vice versa. For the LCC, it might be necessary to make editorial changes, but in general the format should allow for printing the schedules as they are now.

After the Machine-Readable Bibliographic Information (MARBI) Committee of the American Library Association partially reviewed a draft of proposed changes to the authority format to accommodate classification

data, the Network Development and MARC Standards Office decided to rewrite the proposal as a separate format. The office made the decision because it found that there was less overlap with the authority format than originally anticipated, and because the codes and conventions in that format were too constraining. After several revisions, and in further consultation with representatives from the publications of the *LCC*, the *DDC*, the *National Library of Medicine Classification*, and the *Universal Decimal Classification*, the MARBI Committee approved the *USMARC Format for Classification Data* provisionally in June 1990.

The provisional status of the *USMARC Format for Classification Data* means that final approval of the format is delayed until some experience is gained using it; the rules specified in the *USMARC Format: Background and Principles* about the use of content designators will be more flexible than with a fully approved format; experience and experimentation with the format will be of special interest to the Library of Congress and the MARBI Committee; and content designators are more likely to change after institutions experiment with USMARC classification records. The Library of Congress is beginning to experiment with the format for the *LCC* and intends to produce a set of records that other systems may use for experimentation. After sufficient experimentation, the USMARC classification format will undergo further review for possible modifications and final approval.

REVIEW OF USMARC CLASSIFICATION FORMAT

The *USMARC Format for Classification Data* allows for explicit identification of data elements to manipulate the data for a variety of functions.[7] The system itself, that is, the implementation of the format, will largely determine how well those intended functions are met. It is difficult to predict the limitations a system might encounter in using the data; however, the format itself has allowed for great specificity in coding to maintain optimal flexibility and to satisfy likely uses (see figure 1).

Kinds and Types of Records

The *USMARC Format for Classification Data* identifies three kinds of classification records: schedule record (an authority for a classification number from the schedule itself); table record (an authority for a classification number from a table, intended to be added to a base number from the schedule); and index term record (a record for a general explanatory term from a classification index that represents a concept and cannot be associated with one classification number or span). The processing of the record will be largely dependent on the kind of record being coded; for instance, all numbers from the same table will need to be processed and displayed together for comprehensibility. In addition, the type of number, identified as either single number, defined number span (a range of numbers defined by a separate table or subarrangement), or summary number span (a range of numbers that summarizes a topic and is not defined by a separate table or subarrangement) may also determine how the record will be processed. For instance, the fact that the number described in the record is a defined number span indicates the existence of certain other fields defining the table or subarrangement to be applied. These important data elements of the classification number record are contained in the Fixed-Length Data Elements (field 008).

Required Classification Fields

A classification data record must have at the minimum the following fields: 008 (Fixed-Length Data Elements, containing processing information about the contents of the record), 084 (Classification Scheme and Edition,

LEADER	
DIRECTORY	
0XX	Control information, numbers, codes (010-084)
1XX	Classification numbers and terms (153, 154)
2XX	Complex see references (253)
3XX	Complex see also references (353)
4XX	Invalid number tracings (453)
5XX	Valid number tracings (553)
6XX	Note fields (680-685)
70X-75X	Index term fields (700-754)
76X	Number building fields (761-768)

Figure 1. USMARC Classification Tags.

LCC hierarchy:

ITALIAN LITERATURE

Individual authors and works to 1400.

Dante Alighieri, 1265-1321

Translations.

English.

Divina commedia.

Inferno.

PQ4315.25 Particular cantos.

Classification Record:

153 ␣␣‡aPQ ‡a4315 ‡a.25 ‡hItalian literature ‡hIndividual authors ‡hIndividual authors and works to 1400 ‡hDante Alighieri, 1265-1321 ‡hTranslations ‡hEnglish ‡hDivina commedia ‡hInferno

Figure 2. LCC Classification Hierarchy in USMARC Record.

DDC hierarchy:

000 Generalities

001 Knowledge

.1 Intellectual life

Classification Record:

Figure 3. DDC Classification Hierarchy in USMARC Record.

which identifies the classification scheme and edition described by the record), and 153 (Classification Number, which includes the number for which the record is an authority, the caption that describes the subject represented by the number, and the superordinate levels of the caption hierarchy), or field 154 (General Explanatory Index Term, which is a concept not associated with one classification number or span). If appropriate, classification data records may also contain tracing or reference fields, various types of note fields, index term fields (including links to controlled subject headings), and several fields with information on number building.

Treatment of Hierarchies

Several mechanisms are available to communicate hierarchical relationships within the classification scheme. The format accommodates both LCC, which is an enumerative classification scheme, and DDC, which generally uses a hierarchical notation (see figures 2 and 3). Field 153 (Classification Number) contains not only the classification number or span and the caption, but also contains the superordinate caption hierarchy, i.e., all captions to which the caption describing the number(s) is subordinate. This structure not only gives a context for the classification caption, which in some cases may be meaningless (e.g., General works, Study and teaching), but also provides for a hierarchical display (often in the form of indentions) either in an online system or a printed product. By counting the number of subfield ǂh's (caption hierarchy subfield) a classification system could then calculate the indention level of the caption; the caption itself is in a different coded subfield ǂj. In addition, the format provides for tracings to higher or lower captions in the hierarchy to facilitate online browsing; this is accomplished in field 553 (Valid Number Tracing), subfield ǂw, which may be coded to show that the number and caption refer to a broader or narrower topic. The potential for providing links up and down the hierarchy using this special type of tracing is dependent on the capabilities of the systems used for implementation.

Tracings

The classification format includes tracing fields similar to those in the *USMARC Format for Authority Data*. These fields are used to direct the user to another number in cases where the given number is not valid for use or where a different number needs to be considered for classifying the topic. A simple cross-reference display is generated from a tracing field (see figure 4). The tracing field itself is located as a field 453 (Invalid Number Tracing) or 553 (Valid Number Tracing) in the record for the number to which it refers. Tracings in the USMARC authority format work essentially the same way for name and subject authority records and have proven to be very powerful retrieval tools. The format also allows for Complex See References (field 253) and Complex See Also References (field 353) when more detailed instruction is required to convey the information; in these cases the reference leads the user *away* from the number in the classification number field (field 153) to that in field 253 or 353, in contrast to a tracing in fields 453 or 553, which leads *to* the number in field 153 (see figure 5).

Notes

Several note fields may be used to instruct the classifier about the use and application of classification numbers (see figure 6). The Scope Note (field 680) explains topics classed in the number. The Classification Example Tracing Note (field 681) documents the use of a number in another record to facilitate updating of fields when a change is made to the classification number. The Application Instruction Note (field 683) instructs the classifier on the application of tables, subarrangements, etc., sometimes for a particular institution. The Auxiliary Instruction Note (field 684) gives information from a section of a classifier's manual. It is primarily intended for the *DDC Manual*, which in the twentieth edition is published at the end of the *Dewey Decimal Classification* schedule, and is intended to describe policy and practices.[8]

Dewey Decimal Classification users may take particular interest in the History Note (field 685) because of that scheme's ongoing revision and relocation process. This field may be used both for the guidance of classifiers and for computer processing of records to link between old and new numbers. Among other data, the 685 field allows for indicating the type of change recorded, new or previous classification numbers, implementation

LCC reference in schedule:

Economic history and conditions

Agriculture

Agricultural classes

HD(1502) Landlord

see HD1330-HD1331

Classification Record:

153 ♭♭‡aHD ‡a1330 ‡cHD ‡c1331 ‡hEconomic history and conditions ‡hLand use ‡hLand tenure ‡hPolicy. Theory of distribution of the land ‡jLandlord and peasant

453 00‡wj ‡aHD ‡a1502 ‡hEconomic history and conditions ‡hAgriculture ‡hAgricultural classes ‡jLandlord

Figure 4. LCC Tracing.

DDC reference in schedule:

303

.6 Conflict

Class conflict in a specific area of social relations with the subject in sociology e.g. racial conflict 305.8; conflicts as historical events in 900, e.g., the disturbances of May-June 1968 centered in Paris 944.0836

Classification Record:

084 0♭‡addc ‡c20 [Dewey Decimal Classification]

153 ♭♭‡a303.6 ‡hSocial sciences ‡kSpecific topics in sociology and anthropology ‡hSocial processes ‡jConflict

253 2♭ ‡iClass conflict in a specific area of social relations with the subject in sociology, e.g. racial conflict ‡a305.8; ‡iconflicts as historical events in ‡a900, ‡ie.g., the disturbances of May-June 1968 centered in Paris ‡a944.0836

Figure 5. DDC Complex See *Reference.*

6XX Note Fields

680 Scope Note

681 Classification Example Tracing Note

683 Application Instruction Note

684 Auxiliary Instruction Note

685 History Note

Figure 6. Note Fields.

dates, and institution to which the information applies. A tracing field may be used to generate a reference to another number by using a code to indicate whether it is a new or previous number. The tracing does not include as much information as the History Note field and may be used instead of, or in addition to, the note field.

Index Terms

The USMARC classification format includes a block of fields for recording index terms. These fields (700–754) are intended to supplement terms contained in data within the USMARC record itself for additional subject access to the classification number (see figure 7). Fields 700–751 and 754 contain subject access terms controlled by a subject heading system or thesaurus such as the *Library of Congress Subject Headings* (*LCSH*) or *Medical Subject Headings* (*MeSH*). These fields may allow for linking to a subject authority file or to subject headings in bibliographic records. Field 753 (Index Term—Uncontrolled) contains uncontrolled index terms, particularly terms in a back-of-book index to a classification schedule. Separately identified subfields are used in that field to establish hierarchical relationships between terms or for reference to other terms within the classification index. Because the terminology in the classification captions, indexes, and subject headings may all be different within one classification record, subject access is enhanced by providing both controlled and uncontrolled index terms.

The importance of index terms to classification schedules varies from one scheme to the next. Even within *LCC*, treatment of indexes is inconsistent and incomplete.[9] In the current printed LCC schedules, some schedules may have no indexes at all; others may prove indispensible for access to classification numbers. The *DDC*'s Relative Index has an index that "relates subjects to disciplines,"[10] then lists DDC numbers used. In some cases synthesized numbers may be documented only in the index, whereas in the schedules one must apply add instructions to formulate the number. On the other hand, the *National Library of Medicine Classification*'s index links Medical Subject Headings to classification numbers within the scheme and provides guidance on classifying topics.

Number Building Fields

Fields 761–768, the Number Building Fields, provide instructions for the classifier in building classification numbers from sources within the schedule and tables. The fields are heavily coded and are intended for use in computer processing, particularly computer-assisted classification. With the proper system in place, these fields may make it possible for the computer to perform the necessary computations to create synthesized numbers, rather than the classifier. It is largely because of the unusual structure of the 7XX fields that the USMARC Classification Format has been designated a provisional one.

Field 761 (Add or Divide Like Instructions) contains instructions for adding numbers from other parts of the schedule or tables, resulting in a synthesized number. It also identifies a table to be added onto a base number if appropriate. Field 763 (Internal Subarrangement or Add Table Entry) is used to specify an internal classification subarrangement. Experimentation is needed to monitor this field to see how a system can manipulate and display its information; it may be desirable to code each internal subarrangement as if it were an external table.

Field 765 (Synthesized Number Components) may be used to trace components of a synthesized number, showing how the number is formed and where the add instructions appear. It facilitates computer manipulation of synthesized numbers to indicate separately parts of numbers for which it may be desirable to search explicitly, and for computer validation of correct number building. Its usefulness depends on the idiosyncracies of the particular classification scheme; for instance, it will be very useful for the DDC, which relies

70X-75X INDEX TERM FIELDS

CLASSIFICATION FORMAT

700 Index Term–Personal Name

710 Index Term–Corporate Name

711 Index Term–Meeting Name

730 Index Term–Uniform Title

750 Index Term–Topical

751 Index Term–Geographic Name

753 Index Term–Uncontrolled

754 Index Term–Faceted Topical Terms

6XX SUBJECT ACCESS FIELDS

BIBLIOGRAPHIC FORMAT

600 Subject Added Entry–Personal Name

610 Subject Added Entry–Corporate Name

611 Subject Added Entry–Meeting Name

630 Subject Added Entry–Uniform Title

650 Subject Added Entry–Topical Term

651 Subject Added Entry–Geographic Name

653 Index Term–Uncontrolled

654 Subject Added Entry–Faceted Topical Terms

Figure 7. Classification Index Fields in Co with Bibliographic Subject Access Fields.

on number building to create specific classification numbers.

Field 768 (Citation and Precedence Order Instructions) is used for the DDC and contains information about the order in which elements are to be applied in synthesizing a classification number. It also instructs the classifier on order of precedence for the classification of different aspects of a topic.

Sequencing Subfields

Because the note fields and the number-building fields may rely on correct order for interpretation of the data, the Link and sequence number subfield (subfield ǂ8) is provided for linking and sequencing fields. Using this subfield when necessary, the inputter can indicate the order in which fields must remain to provide comprehensible instructions. This ability is particularly important for the *DDC* because of the editorial style rules that specify the order in which notes should appear in the printed schedules and tables. In addition, both *LCC* and *DDC* include internal tables and subarrangements of data that will be input into separate fields in USMARC records; the comprehensibility of the instructions relies on the fields remaining in a certain order. Subfield 8 will ensure that the fields are processed in the correct order in these situations (see figure 8 for an example of the use of the sequencing subfields).

IMPLEMENTATION OF USMARC CLASSIFICATION

Whether the *USMARC Format for Classification Data* is effectively utilized as an online tool for various functions largely depends on how it is implemented. The online system that will be used to process, display, maintain, and print USMARC classification data has not yet evolved. Institutions implementing USMARC classification need to write detailed specifications for systems to make use of the records. Most important will be the design of a browse display so that users can see classification data from one record in relation to that from other records. The individual record approach alone will not suffice for representing classification information; it needs to be seen in its hierarchical context of subject terms. Most problematic may be the ability of a system to provide a comprehensible display of classification tables.

The Library of Congress is beginning to explore the required conversion of the *LCC* from forty-six printed classification schedules to online records in a USMARC database. Before the project can commence, the schedules will need manual editing to generally reduce inconsistencies from one schedule to the next, to verify classification numbers in notes and references, and to eliminate many of the space-saving practices, such as foot-

Internal table in LCC schedule:

389 By region or country, A-Z

Under each country:

.x General works

.x2 Inland

.x3 Ports

.x4 Special waterways, rivers whose course is entirely within the limits of one country, A-Z

Classification record:

153 ##‡aHE ‡a389 ‡a.A ‡cHE ‡c389 ‡c.Z ‡hTransportation and communication ‡hWater transportation ‡hWaterways ‡hControl, taxation, tolls, etc.‡jBy region or country

763 8#‡81.1 ‡iUnder each country:

763 10‡81.2 ‡a.x ‡jGeneral works ‡p153

763 10‡81.3 ‡a.x2 ‡jInland ‡p153

763 10‡81.4 ‡a.x3 ‡jPorts ‡p153

763 11‡81.5 ‡a.x4 ‡aA ‡c.x4 ‡cZ ‡jSpecial waterways, rivers whose course is entirely within the limits of one country ‡p153

Figure 8. Internal Table in LCC.

notes, "divide like" notes, or use of the plus sign in references.

The process of inputting records from all of the *LCC* schedules seems an overwhelming prospect. It is estimated that an *LCC* database will consist of approximately 450,000 to 500,000 records.[11] Computer experts from the Library's Cataloging Distribution Service have reviewed the USMARC classification format and concluded that the format could support the printing of *LCC* schedules as they now appear if the system were in place.

Currently, the Library of Congress is conducting an experiment to evaluate the new USMARC format and to consider the editing process that *LCC* needs to undergo to implement it. The Network Development and MARC Standards Office, in conjunction with the Office for Subject Cataloging Policy and the Subject Cataloging Division, are overseeing the encoding of classification data from *LCC*'s H (Social Sciences) schedule. Records for classification numbers in schedules H through HG have been input into a system for the creation of USMARC records on a microcomputer using the Minaret software. Because the individual record approach does not adequately represent classification information that needs to be viewed in its hierarchical context of subject terms, the developer of the software has prepared an enhancement for LCC to enable the user to access a classification browse display. This display is generated from an enhanced index, and it displays the data on the screen in a format similar to the page of a classification schedule. Features of

the Minaret classification database include the following:

- Hierarchy of captions are shown as indentions, as in the printed *LCC*;
- Reference notes are generated from tracings and appear in their proper place, with display text generated;
- References to other sections of the LCC schedules outside of H–HG are input as tracings in records coded as "incomplete";
- Invalid numbers are identified by an asterisk (*);
- Numbers have been assigned to all captions in the printed schedules (most numbers that do not appear in the printed schedules, but are needed in the online record, are for summary number spans or invalid numbers);
- When a table is to be applied to a span of numbers, it can be windowed onto the screen;
- Some particularly complex internal tables applied to many numbers have been coded as external tables;
- Not all notes in the full record appear in the browse display; instead, the word *note* is generated, to alert the user to view the full record;
- Separate indexes of captions and back-of-the-book index terms are available; also, Boolean searches may be performed;
- Update of records is easy and quick.

The section of the Subject Cataloging Division that classifies material in the H schedules is using the classification database in its ongoing work to classify current materials. It is expected that the experiment will continue to include other portions of the *LCC*. The records being created within this project may later be incorporated into a larger USMARC-based classification system when it is in place. In addition, the Library of Congress is considering ways to make this set of USMARC classification records available to other institutions for experimentation.

The Decimal Classification Division plans to supply the *Dewey Decimal Classification* in USMARC format in the future. It is exploring the conversion of records from the Editorial Support System (ESS) format, which was developed to publish the twentieth edition of the *Dewey Decimal Classification*, into the USMARC format. The division plans to continue using the ESS format for editing and printing the classification, and also expects to use an automatic conversion program to produce USMARC records for distribution. Changes are being made in the ESS database as work has begun on the twenty-first edition, scheduled for publication in 1996.

CONCLUSIONS

Now that a standard USMARC format has been developed for classification data, there is much work left to do to make online classification data a reality. "Standards are both the blessing and the bane of libraries. For, while enormous benefits are reaped when standards are agreed upon and applied, considerable consequences are wrought when, having been established, standards are neglected or changed. However difficult they are to attain or expensive to implement, they are absolutely critical to the usefulness of bibliographic data, especially in large databases."[12]

Whether the USMARC classification format lives up to its potential is highly dependent on the ability of online systems to use the wealth of data in classification numbers and captions effectively and efficiently. The design of future retrieval systems for classification records may determine whether the USMARC classification format becomes the powerful tool for subject access, maintenance of classification schedules, and machine-assisted classification that it has the potential to become.

REFERENCES

1. Lois Mai Chan, "The Library of Congress Classification System in an Online Environment," *Cataloging & Classification Quarterly* 11, no.1:9 (1990).
2. John J. Finni and Peter J. Paulson, "The Dewey Decimal Classification Enters the Computer Age: Developing the DDC Database and Editorial Support System," *International Cataloging* 16, no.4.
3. Karen Markey and Ann Demeyer, *Dewey Decimal Classification Online Project: Evaluation of a Library Schedule and Index Integrated into the Subject Searching Capabilities of an Online Catalog: Final Report to the Council on Library Resources* (Dublin, Ohio : OCLC Online Computer Library Center, 1986).

4. Lois Mai Chan, "Library of Congress Classification as an Online Retrieval Tool: Potentials and Limitations," *Information Technology and Libraries* 5, no.3 (Sept. 1986).
5. *The USMARC Formats: Background and Principles* (Washington, D.C.: Network Development & MARC Standards Office, 1989), p.4.
6. Walt Crawford, *MARC for Library Use*, 2d ed. (Boston: Hall, 1989), p.7.
7. *USMARC Format for Classification Data: Including Guidelines for Content Designation* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1991).
8. *Dewey Decimal Classification and Relative Index*, v.4 (Albany, N.Y.: Forest Pr., 1989), p.731–958.
9. John Phillip Immroth, *Analysis of Vocabulary Control in Library of Congress Classification and Subject Headings* (Littleton, Colo. : Libraries Unlimited, 1971), p.82.
10. *Dewey Decimal Classification and Relative Index*, v.1 (Albany, N.Y. : Forest Pr., 1989), p.xliv.
11. Nancy Williamson, memorandum to author, Mar. 19, 1991. Includes general results of her study on the *Library of Congress Classification*.
12. Henriette D. Avram, "Building a National Information System: The U.S. Perspective," unpublished; presented at "National Library in the Life of the Nation," jointly sponsored by the Lenin State Library and the Library of Congress in October 1991. ■■

Communications

Rates and Types of Changes to LC Authority Files

Karen Calhoun and Mike Oskins

This paper presents new findings on how the Library of Congress (LC) authority files change over time. With the refinement of vendor services for one-time automated authority processing and local system authority control modules, there has been increasing interest in methods for keeping local system files up-to-date following an initial, costly database preparation.[1,2] However, there has been no research published on how the LC authority files change and how the changes might impact local databases.

The authors investigated authority record transactions issued by LC to provide reliable data on:

1. The number and percentages of new, changed, and deleted name and subject authority records being issued by LC;
2. The percentage of changes affecting the authorized heading (lxx) field;
3. The daily rates of change of (a) heading fields and (b) all fields; and
4. The rate of occurrence of multiple changes to the same record over a thirty-day period.

Because the data can be used to build a model of how LC headings in local system databases age, the findings of this study will be of interest to library technical services or database maintenance staff, to managers who need to make cost estimates or plan work flows, and to systems staff and vendors who provide products and services based on the LC authority files.

METHODOLOGY

The authors based the analyses on updates to the LC Name Authority File (NAF) and LC Subject Authority File (SAF) that LC issued over thirty production days in spring 1991. The updates consisted of new authority records, changes to existing records, and deletions. OCLC receives updates to the LC NAF from LC daily via the Linked Systems Project (LSP) Authorities Implementation; updates to the LC SAF are received weekly on tape.

The initial data file included all new records, changes, and delete transactions on NAF and SAF records over the thirty-day period. However, the only analyses of new records and delete transactions were simple frequency counts. The study concentrated primarily on change transactions—that is, changes to already existing LC authority records in the NAF and SAF.

As a first step, a program removed Change Message Records (CMRs)—temporary records that are used by LC to alert catalogers that a change to a name authority record is in progress—from the data file of change transactions. Easily identified, CMRs contain code "b" in a fixed field element (008/31, Record update in process). The authors chose to exclude CMRs from the study because they exist only in LSP systems and because, even if CMRs were generally available, they are irrelevant for database maintenance.

After the removal of CMRs, software developed by the OCLC Office of Research created pre- and postimage authority records for each change transaction. The pre-image record stored the record as it was prior to the change, and the postimage record stored the changed record. Next, for each pre- and postimage pair, comparison software created field change records for every field added or modified. The field change records were then input to several programs for analysis and printing.

Because changes to authorized heading (lxx) fields have the greatest impact on bibliographic databases, the authors examined them extensively. The software selected the sample of heading changes by outputting change records whose field tags began with

Karen Calhoun is Manager, Online Data Quality Control Section and Mike Oskins is Consulting Systems Analyst, Office of Research, OCLC.

Table 1. Number and Percentage of New, Changed, and Deleted Records

	Average No. Transactions/Day			% Transactions/Day			Totals		
Transaction Type	NAF	SAF	Both	NAF	SAF	Both	NAF	SAF	Both
New records	807	19	826	60	38	60	24,219	571	24,790
Changed records	506	30	536	38	60	39	15,181	887	16,068
Deletes	25	1	26	2	2	2	758	26	784
Totals	1,338	50	1,388	100	100	101*	40,158	1,484	41,642

* Rounding error

"1," the digit used for all authorized heading fields in authority records.

The last analysis was a longitudinal study, which was accomplished by separating out pre- and postimage records that reappeared two or more times during the thirty-day period. The comparison software then analyzed the changes in the multiple-occurring record file.

FINDINGS

New Records, Changes, and Deletions

Table 1 gives counts and percentages of new records, changes, and deletions of name and subject authority record updates over the thirty production days. Because the data were collected from OCLC daily journal files, results are reported in transactions per day, even though subject updates are processed on a weekly basis.

Besides the great difference in transaction volume for NAF versus SAF updates (for every SAF update, there were more than twenty-five NAF updates), note that for names, the majority of updates were new records (60%), whereas for subjects the majority of updates were changes to existing records (60%).

Change Transactions

The 16,068 change transactions were selected for additional analyses. The change transactions contained a total of 27,559 added or modified variable fields, for an average of 1.7 added or modified fields per record. Added fields accounted for 55.6% of the changes, and modified fields accounted for 44.4%.

As indicated earlier, the software discarded CMRs prior to the full analysis. After removing CMRs, 14,793 (92.1%) of the original 16,068 change transactions remained.

Heading (1xx) Field Changes

Of the remaining 14,793 change transactions, 4,732 affected the authorized heading (1xx) field. Table 2 breaks down the heading field changes by tag. More than half of the changes to heading fields were changes to personal name headings (field 100); about one in five heading changes were to corporate names (field 110).

Scanning selected heading change pairs quickly revealed that many heading changes were the result of a simple removal of the terminal period in the heading field, for example

pre-image 100 10 Gawler, Ian.
postimage 100 10 Gawler, Ian

About two years ago, LC issued instructions to NAF contributors that were intended to bring the punctuation conventions for NAF records into line with those for SAF records. The instructions directed catalogers to remove the terminal period from the NAF record heading field if the record was being changed for any reason. The findings of the present study confirm that NAF contributors are following LC's directive. In fact, 2,657 of the 4,732 heading field changes (56.1%) involved the removal of the terminal period.

Table 3 gives a breakdown of the 12,136 changes that were left after excluding the CMRs and insignificant heading changes consisting solely of terminal punctuation removal.

Effect of Normalization

Two additional analyses were performed on the remaining 2,075 heading changes to study the effect of text normalization routines

Table 2. Heading Field Changes by Tag

Field Tag	No.	%
100 (personal name)	2,774	58.6
110 (corporate name)	1,117	23.6
111 (conference name)	173	3.7
130 (uniform title)	415	8.8
150 (topical term)	154	3.3
151 (geographic name)	99	2.1
Totals	4,732	100.0

Table 3. Frequency of Heading Field Changes (Excluding Terminal Punc. Removals)

Transaction Type	No.	%
Heading (1xx) changes	2,075	17.1
Changes to other fields	10,061	82.9
Totals	12,136	100.0

(beyond the removal of terminal periods). The first normalization routine converted all letters to uppercase, removed diacritics and marks of punctuation, and stripped content designators (i.e., indicators, subfield delimiters, and subfield codes) before input to the comparison software. In this analysis, another 391 (18.8%) of the heading field changes were removed, leaving 1,684 heading changes (13.9% of the 12,136 change transactions).

A second normalization routine retained differences resulting from content designator (i.e., MARC coding) changes but removed differences resulting from changes to case, presence or absence of diacritics, and changes in punctuation. In this analysis only 288 (13.9%) heading field changes dropped out, leaving 1,787 heading changes (14.7% of the 12,136 changes).

Table 4 gives some examples of heading change pairs that would or would not be considered a heading change under the two normalization routines. From a database management perspective, the second normalization routine is probably superior, because it removes insignificant changes without eliminating valuable information that can be used to control sorting and display of headings in library catalogs.

To summarize, terminal period removals and the less strict normalization routine eliminated 2,945 (62.2%) of the original 4,732 heading change transactions. Terminal period removals and the strict normalization routine eliminated 3,048 (64.4%) of the original 4,732 heading change transactions.

Rates of Change

All Changes

Excluding CMRs and heading field changes that consisted solely of the removal of terminal punctuation, there were 12,136 change transactions over the study period, an average of 405 changes per production day.

Heading Changes

Excluding heading changes that involve only the removal of a terminal period and applying the normalization scheme that retains content designators, there were 1,787 heading field changes over the study period, an average of 60 heading changes per production day. The study data indicate that slightly more than one in ten of the changes being issued by LC is a change to the authorized form of the heading (1,787 of the original 16,068 change transactions).

Multiple Occurrences of the Same Record

The longitudinal analysis separated out the pre- and postimage pairs that appeared more than once during the thirty-day period. After removing single occurrences of a pre- and postimage pair and multiple occurrences involving the distribution of a CMR and subse-

Table 4. Effect of Normalization Routines (Excluding Terminal Punc. Removals)

Example				Considered a Changed Heading? (Y/N)	
				Strict Normalization	Retain Content Designators
Post	110	20	Colonel Mike's Dance Band	N	Y
Pre	110	10	Colonel Mike's Dance Band		
Post	130	b0	Coleccion Biblioteca popular	N	N
Pre	130	b0	Coleccion biblioteca popular		
Post	100	10	Britt, John, $d 1927–	Y	Y
Pre	100	10	Britt, John		
Post	100	10	Bernal, Luis Darío	N	N
Pre	100	10	Bernal, Luis Darío		
Post	130	b0	OPm (Ames, Iowa)	Y	Y
Pre	130	b0	OPm		
Post	100	10	Pratt, Vince E., $d 1912–	N	Y
Pre	100	10	Pratt, Vince E., 1912–		

quent changes, 2,988 (18.6%) of the original 16,068 change transactions remained—that is, were found to be multiple change transactions on the same record during the thirty-day period.

The great majority of repetitions of the same record occurred twice during the study period; some repeated three times. One record repeated five times.

FUTURE RESEARCH

The study demonstrates the dynamic nature of the LC NAF and SAF. It is safe to assert that if it is not kept up to date in some manner, inaccurate and conflicting headings will accumulate rather rapidly in a local system file, even one which has been cleaned up and brought into conformance with the LC NAF and SAF before being loaded.

The findings also suggest a type of change—terminal period removals on lxx heading fields—that could be accomplished via a database scan of the LC NAF, rather than via manual efforts of NAF contributors.

Future research might examine:

1. The types of nonheading field changes being made. Comparison software could be used to provide more detailed information about what LC and other contributing libraries are doing to nonheading fields in authority records. Such analyses could be used to answer questions such as:

- Are there other changes (besides the removal of terminal punctuation) now being done manually that could be done via database scan techniques?
- What are the number and types of nonheading field changes? Are there patterns of change?
- What are the most important nonheading changes? Which ones have the greatest impact on local databases and how? Are there changes that are insignificant from a database maintenance or public services perspective?

2. How name authority updates differ from subject authority updates.

3. The impact of authorized heading field changes on a bibliographic database.

The study data indicate that heading changes for personal names (100 fields) are the most voluminous. However, although topical heading field (150) changes are fewer in number, could they nevertheless have a greater impact on a bibliographic file?

REFERENCES

1. Elaine Peterson and Bonnie Johnson, "Is Authority Updating Worth the Price?" *Technicalities* 10, no.5:1–15 (1990).
2. Alva Theresa Stone, "Vendor Processing and Local Authority File Development," *Law Library Journal* 81, no.1:131–41 (1989). ■■

Reviewing Initial Stopword Selection

Bonnie Johnson and Elaine Peterson

Five years ago stopwords were selected for our online catalog based on conventional wisdom and analysis of the database as it looked then. It can be documented that a database can sufficiently change in just a few years, making it necessary to reevaluate the stopword list initially selected.

A standard definition of a stoplist might be a list of words or terms, or roots of words, that are considered to be meaningless or insignificant for purposes of information retrieval and that are excluded from indexing. Therefore, the purpose of stopwords is to make indexing and searching of a database more efficient. By registering basic words such as *the* or *and* as stopwords, one is able to avoid the time-consuming (or CPU-consuming) task of indexing those words in the first place and to stop their retrieval if entered as a keyword search by a user.

Table 1. Original Soft Stopwords

American	Dept
Annual	Institute
Association	Institution
Bulletin	International
Bureau	Introduction
Committee	Journal
Conference	Meeting
Company	National
Congress	Report
Council	Society
Department	Symposium

Bonnie Johnson is Head, Collection Maintenance, and Elaine Peterson is Assistant Dean for Technical Services, Montana State University Libraries, Bozeman.

When our database was being profiled and built in 1987, a task force studied the problem of determining which words to enter as stopwords. Fortunately, our online catalog (which runs on Inlex software) is one that allows for both "hard" and "soft" stopwords.

Table 2. Original Soft Stopword Occurrence

Original Soft Stopwords	1991 Occurrence
Institution	458
Department	641
Company	901
Meeting	968
Bureau	1,016
Bulletin	1,192
Annual	1,364
Congress	1,792
Dept	1,826
Council	2,021
Journal	2,507
Committee	3,157
Association	3,478
Institute	3,656
Introduction	3,941
Symposium	5,023
Report	5,045
Conference	5,602
Society	5,937
National	6,659
International	7,663
American	14,397

Table 3. All Occurrences over 4,000 times

Word	1991 Occurrences	Word	1991 Occurrences
States	25,856	Literature	5,844
United	25,177	Theory	5,707
History	22,287	Proceedings	5,691
American	14,397	James	5,686
Congresses	12,315	Life	5,686
English	11,274	Conference*	5,602
Montana	10,832	State	5,493
John	10,358	World	5,376
Research	9,666	Guide	5,369
Education	9,554	Aspects	5,326
New	9,014	Management	5,239
Social	8,478	Great	5,130
Series	8,273	Analysis	5,076
Study	8,222	Report*	5,045
University	7,928	Symposium*	5,023
William	7,715	Canada	5,013
Periodicals	7,693	Water	4,635
International*	7,663	David	4,602
Studies	7,605	School	4,561
Robert	6,922	Modern	4,553
America	6,741	Public	4,531
National*	6,659	Engineering	4,374
Science	6,587	Policy	4,308
Century	6,585	Health	4,260
Art	6,149	Charles	4,248
Development	6,146	Works	4,203
Bibliography	6,097	Richard	4,188
Economic	6,090	George	4,166
Society*	5,937		

* Original stopword

(Many online catalogs do not make this distinction, but most allow for adjustment of stopwords.) Hard stopwords are those entered that can never be changed, while soft stopwords can be adjusted as needed. All parts of speech except nouns were included as hard stopwords: articles (*the, a, an*), prepositions (*by, in, on*), pronouns (*he, she, it*), verbs (*has, are, been*), adverbs (*as, how*), adjectives (*my, their*), conjunctions (*and, or, not*), and abbreviations (*etc, inc*).

A more lengthy discussion was needed for the creation of the soft stopword list. In the end, based on preliminary analysis of the database as it looked then and conventional advice about stopwords, the words in table 1 were entered as soft stopwords. This list includes nouns.

Table 4. Increase in Stopword Occurrence

Word	% Increase 5/88–5/91	Word	% Increase 5/88–5/91
Engineering	116	James	29
School	80	Policy	27
Developments	69	Theory	25
Aspects	55	Great	24
National*	52	Proceedings	24
Modern	51	David	23
Science	49	State	23
Works	44	Economic	22
Montana	43	History	22
Report*	43	Richard	22
Health	40	Social	22
New	39	Study	22
World	38	Robert	21
Art	37	Guide	19
International*	37	Literature	19
Public	36	Periodicals	19
Water	36	States	18
Analysis	35	Studies	18
Society*	35	United	18
Education	34	Conference*	17
George	34	English	17
Management	33	Life	17
William	33	Symposium*	17
John	32	Series	16
Century	31	America	14
Charles	31	Congresses	14
Canada	30	Research	13

* Original stopword

When we received a major new software release, we decided to reindex our entire database. It was, therefore, an appropriate time for evaluation of potential candidates for stopwords. A report was produced showing the number of items attached to each word in the original "soft" stopword list (included in table 2; totals in this and subsequent tables are for author, title/series, and subject fields). It is readily apparent that some of the selected stopwords have had few occurrences and little impact on efficient searching.

As a point of comparison, a report was run in 1991 that listed all words with more than 4,000 occurrences (see table 3). The initially selected soft stopwords only account for six of the fifty-seven words that now occur more than 4,000 times.

Even more interesting is an examination of the percentage increase over the last three years of those words that occur more than 4,000 times in the database (see table 4). Given the scope of our collection development policy, the percentage increase of words like *engineering, school, Montana, health,* and *art* is no surprise. We would not be able to make most of these high-occurence words into stopwords because they are meaningful

words to search, even if their occurrence is high. However, there are types of words similar to those we originally indexed that, given their rate of increase, make sense as stopwords. Words like *works, new,* and *proceedings* are increasing faster than most of the words we entered originally as stopwords. It makes as much sense to enter these, or a word like *aspects*, which has increased by 55 percent in just three years, as it did initially to enter *meeting* or *annual*, which have increased less than 2 percent. The high occurrence and rate of increase of surnames like *George, William,* and *John* make these candidates for stopwords. In all cases, it is important to ascertain whether patrons are likely to search on a given word. Although some words may appear to be unimportant, they may in fact be integral to certain titles, such as the word *works* in a book about the Works Progress Administration. Both *John* and *Montana* are high-occurrence words also, but whereas we may want to block users from keyword searching of *John*, we would not for a word like *Montana.* It is also important that a message appear to patrons explaining that they have searched on a stopword and must enter a different term or delete their initial choice from the search.

In conclusion, based on our experience, words that are initially perceived to be candidates for inclusion in a stopword list can be quite different than those that are truly increasing in a database. The rate of change of any given database over time has led us to recommend that stopwords be examined every few years. Patrons should be able to retrieve satisfactory results via keyword searches without placing unnecessary strain on the computer system. ■■

Prediction of OPAC Spelling Errors through a Keyword Inventory

Terry Ballard and Arthur Lifshin

In order to find and correct spelling errors in the online public access catalog at Adelphi University, a visual inspection was performed of the 117,000 keywords indexed in the system. More than 1,000 errors were found. Certain long but common words such as administration, education, *and* commercial *were found to generate many different misspellings. Most of the records were derived from bibliographic utilities, so the findings can be generalized to other OPACs. The same misspellings were also found in substantial numbers in CD-ROM databases. Misspellings were analyzed by the machine-readable catalog (MARC) field in which they were found, part of speech, and type of mistake. Lists of commonly misspelled root words and specific mistakes are included.*

Terry Ballard is Assistant Professor and Systems Librarian, Adelphi University, Garden City, New York, and Arthur Lifshin is Head, Science and Engineering Center, Folger Library, University of Maine at Orono.

In the years since the online public access catalog (OPAC) has replaced the card catalog as the primary source of bibliographic information in research libraries, much has been written about miskeyings of library users by Peters and Blazek, among others.[1,2] However, little attention has been paid to spelling errors that become a part of the database. Perhaps this is because researchers can work with logs of OPAC transactions to find the searching errors of OPAC users, but there is no easy way to get at misspellings that are in a database containing millions of words.

It may be widely perceived that spelling errors in OPACs and other large databases are few in number, randomly distributed, and impossible to locate in any systematic fashion. The results of this study demonstrate that these perceptions are incorrect.

HISTORY OF THE STUDY

In a recent issue of *American Libraries,*[3] there appeared a short article describing how Jeffrey Beall at Harvard had found words that are prone to misspelling, such as *Febuary* or *goverment.* Librarians at Adelphi University Library in Garden City, New York, checked the keyword index in the library's Innovative Interfaces OPAC (Innopac) and found single examples of two of the ten words that were featured. According to a formula provided in the article, Adelphi had a very clean database.

This was not surprising because the cataloging supervisor has described the operation as one with a history of perfectionism.

However, misspelled words did show up occasionally, and the author found a way to make a thorough search of the database for such problems. In the Innopac system, a successful keyword search will display a record that contains the word that was queried or a menu if there is more than one hit. However, if there is no matching word, it will produce a screen of choices that are nearby in the alphabet. One may then browse forward or backward looking at eight titles per screen

A trial search of the *A*'s was performed by typing in *aaaaa* and browsing forward through hundreds of screens. The result was the identification of forty-two spelling errors. This justified a search of the 117,000 words that are contained in Adelphi's 310,000 bibliographic records. A complete visual check of the keyword index represented a large volume of work, but it seemed like the only reasonable way to get the problem solved. Normally, two letters were searched in a single workday. Once a potential spelling mistake was identified, the full record was called up and checked for context, e.g., *langage* is correct in French, but it is a misspelling in English. If it did turn out to be a mistake, the screen containing the incorrect word was printed along with the menu screen of eight words that contained the error. At the end of reviewing a letter, the printout was marked for immediate correction by a student assistant. The system allows the staff member to call up a record, identify the field with the misspelling, and substitute the correct word for the misspelled one.

In going through the screens, one major problem was non-English words that contained only a single letter's difference from the English equivalent. Another snag was archaic spellings that actually appeared in the title pages of old books. Shakespeare was a particular problem—some books published in the twentieth century used the spelling *Shakspeare*. Other records had that spelling even though the book did not. The system does not display punctuation marks, so there were many three-letter initials and Roman numerals that displayed as nonsense words.

After we had completed the inventory, we ran global searches of the database to find any of the commonly misspelled words in the subject fields. This was accomplished in the list creation mode of Innopac. A search could be run for four words at a time using a search of character strings. We did not find any misspellings in the subject fields.

LITERATURE

A search of the literature resulted in a dearth of material on misspellings in an OPAC. Bourne looked at the frequency of spelling errors in bibliographic databases in 1977.[4] Covered in the survey were such databases as ERIC, *Chemical Abstracts Condensates*, BIOSIS, NTIS, EI, NAL, PsychAbst, ISMEC, ABI, and PATS. The percentage error found was 0.29% overall, which ranged from a low of 0.09% to a high of 0.89%.

Cynthia Ryans surveyed OCLC records in 1978 and examined the distribution of errors in fields and the distribution of types of errors.[5] The errors involved in this study included all types of cataloging mistakes. Misspellings were mentioned but not separately analyzed. Of the errors, 58% were in the collation field, 31% in subject headings, 19% in series description, and 16% in both the title and added entries.

A series of articles by Zamora; Zamora, Pollock, and Zamora; and Pollock and Zamora discusses the distribution of spelling errors in *Chemical Abstracts* and the techniques (algorithms) used to clean up the database.[6] Their results showed that, while the number of spelling errors found was large (50,000), when compared with the total number of records the error rate was only 0.2%. They found that omissions were the most common errors (30–40%), followed by insertions (25–35%), substitutions (15–20%), transpositions (10–15%), and multiple errors (4–9%).

O'Neill and Vizine-Goetz looked at the OCLC database spelling errors in the light of possible correction mechanisms.[7] Although they list some sixty-four spelling variants of *criticism*, they do not investigate the actual distribution of errors within the database. They conclude that, although spelling checkers are impractical for very large databases, they can be very useful in the case of smaller databases. They also conclude that error-correction algorithms in conjunction with spelling checkers and good editing techniques are the best available solution.

The various solutions outlined in the above articles are impractical in the case of Adelphi's

OPAC. Due to the nature of the OPAC it is not possible to introduce a spelling correction program into the database. Transferring the entire database out of the OPAC and into another system and then back into the OPAC would be prohibitive in time and cost. Also, it would probably introduce more errors than it would remove. The added problem of the many listings of foreign words and names would, at this juncture, still bring up tens of thousands of the items in our index.

METHODS AND RESULTS

The spelling errors were entered into a database in Paradox 3.5. Data entered included the word or root word, location in the record (title, author, note, etc.), part of speech (noun, verb, etc.), origin (OCLC, LC-MARC, or local), type of error (omission, insertions, etc.), and OCLC number. Table 1 shows the distribution of errors in terms of location, table 2 as part of speech, table 3 as origin, and table 4 as error type.

It should be noted that 60.26% of the spelling errors are found in the title fields, followed at a distance by the notes field at 21.35%; the rest are minor. The same pattern is true for parts of speech, with 60.63% of the errors occurring in nouns, followed by adjectives and verbs at 18.12% and 13.86%, respectively, the rest being minor. Origin is dominated by OCLC records with 85.31%. Error types are omission (49.08%), insertions (19.78%), transpositions (11.01%), and substitutions (12.94%). Dropped and added spaces represented 5.36% and 1.76%, respectively.

With the exception of a slightly higher value for omission, the data for the error types agree with Zamora. The inclusion of added and dropped spaces seemed justified on the basis of occurrence, a total of 7.12%. If these were added into the omissions and insertions, the results would skew still further. Given the dominance of the title field, it is not surprising that nouns also dominate; one would expect book titles to contain primarily nouns.

Since 85.94% of Adelphi's bibliographic records are OCLC, it is not surprising that 85.31% of the errors are found in OCLC records. A 50,000-record sample of OCLC records from the Adelphi database was generated, and the distribution of OCLC numbers by "millions" was compared with the distribution of OCLC numbers in the error group. A linear regression was run between

Table 1. Distribution of Spelling Errors by Field Location

	No.	%
Author	21	1.94
Title	652	60.26
Alternate Author	71	6.56
Alternate Title	32	2.96
Notes	231	21.35
Series	71	6.56
Imprint	4	0.37
Total	1,082	100.00

Table 2. Distribution of Spelling Errors by Part of Speech

	No.	%
Noun	656	60.63
Verb	150	13.86
Adjective	196	18.12
Adverb	58	5.36
Miscellaneous	22	2.33
Total	1,082	100.00

Table 3. Distribution of Spelling Errors by Origin

	No.	%
OCLC	923	85.31
LC-MARC	24	2.22
ADELPHI-CAT	57	5.27
VERT.-FILE	78	7.21
Total	1,082	100.00

Table 4. Distribution of Errors over Error Types

	No.	%
Omission	531	49.08
Insertion	214	19.78
Substitution	140	12.94
Transposition	120	11.01
Added Space	19	1.76
Dropped Space	58	5.36
Total	1,082	100.00

the two with the resulting $r = 0.96$ showing a very high degree of correlation, indicating that the distribution of OCLC numbers is the same for both groups. Figure 1 shows the distribution of OCLC numbers (in millions) for the error group.

A determination was made of the position of the spelling error in the title—whether in the subtitle, first word, second word, etc. Of the 652 title errors, 358 (54.91%) are in the subtitle. The rest are distributed as shown in table 5. This demonstrates that a misspelled word in the title field will most likely be in the subtitle or within the first few words of the title. The distribution shows equal probability of the error being in the first two words; the probability slowly diminishes until the eighth word, where there is a sudden drop. This will undoubtedly have an effect on title searching; the extent to which it does will depend on the number of errors in the database.

It became obvious very early in the study that some words are misspelled more often than others and that words that you expect to be misspelled are not. Variants of the same word were collected as one; endings and common suffixes were ignored in generating the word counts. All of the words that are misspelled many times tend to have eight or more letters and at least three syllables. Table 6 lists the most commonly misspelled words and their frequencies. Although these may not be common in everyday use they are quite commonly found in any academic OPAC. It is the more common words that have been misspelled and not the more esoteric technical terms.

Table 5. Distribution of Errors within the Title Field

Word Number	Amt.	%	Cum. %
1	42	14.29	14.29
2	42	14.29	28.58
3	36	12.25	40.83
4	32	10.86	51.70
5	29	9.86	61.57
6	25	8.50	70.05
7	28	9.52	79.57
8	14	4.76	84.33
9	7	2.38	86.71
10	8	2.72	89.43
>10	31	10.54	100.00

Words such as *development, psychology,* or *possibility* would show many different permutations of dropped or inverted letters. In the case of *psychology,* we started encountering these with *paychology* and kept seeing variations through *pyschology.* Table 7 shows some of the more common misspellings that occurred in the Adelphi OPAC.

There were several possible mistakes that generated specific searches during the project. We thought that common words such as *the* or *and* might show up in combinations such as *teh* or *adn.* They did not. Another possibility was words that ended in some combination of *tion* showing up as *toin.* A search was run of the database for any word in the title or notes field containing the string toin. More than fifty were called up, but none was a misspelling.

It must be emphasized that the errors in our OCLC records did not necessarily originate with OCLC. Using our file of OCLC error numbers sorted in numerical order, we selected every tenth record to retrieve. We looked at them to determine whether mistakes were still present. In doing so we found that some mistakes were in subtitles that did not exist in the OCLC record, so we surmise that the spelling error was input during local editing. However, of the 123 records that we called up, 63 still had a misspelled word in the OCLC database.

DISCUSSION

Since most of the misspelled words were on records that were derived from national databases, it seemed reasonable to suspect that these commonly misspelled words would be present in other OPACs. The above words were searched at a number of other large libraries in the New York metropolitan area, and it was verified that these problems are widespread. In addition, we looked at some of the more common misspellings in the CD-ROM databases that we have and found hundreds of hits on words such as *commerical, reseach, adminstration,* and *government.*

RECOMMENDATIONS

Obviously, every library that has an OPAC with keyword capability should search the problem words that we have identified and fix the inevitable errors. Even when the mistakes are corrected in OCLC, member libraries using the records prior to correction will still

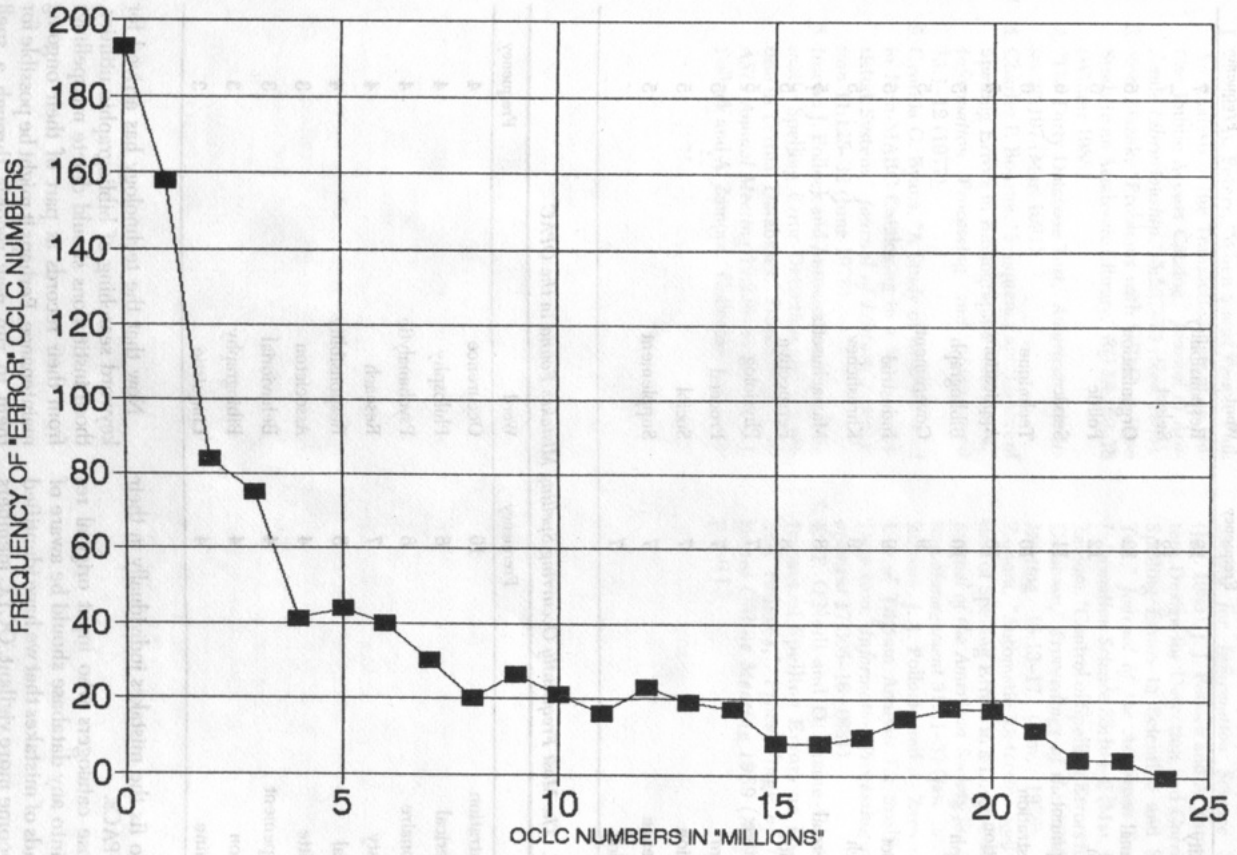


Figure 1. Distribution of "Error" OCLC Numbers

Table 6. Most Common Misspelled Word Roots Found in the OPAC

Word	Frequency	Word	Frequency
University	18	Responsibility	7
Histor	12	Select	7
Psychoanal	12	Organization	6
Relation	12	Politic	6
Development	11	Servic	6
Administration	10	Technique	6
Association	10	Applications	5
Philosoph	10	Bibliograph	5
Integ	9	Government	5
Introduct	9	Industrial	5
Research	9	Khrushchev	5
Commercial	8	Massachusetts	5
Education	8	Perspective	5
Committee	7	Physiolog	5
Criticism	7	Proceed	5
Institution	7	Sociol	5
Occurrence	7	Supplement	5
Psycholog	7		

Table 7. The Most Frequently Occurring Spelling Mistakes Found in the OPAC

Word	Frequency	Word	Frequency
Adminstration	10	Occurence	4
Commerical	8	Philosphy	4
Questionaire	8	Psychanalytic	4
Universty	7	Reseach	4
Intergral	5	Responsiblity	4
Committe	4	Associaton	3
Developement	4	Behaviorial	3
Educaton	4	Bibiography	3
Pantomine	4	Linquistic	3

need to fix the mistakes individually in their own OPACs.

Those catalogers who input original records into any database should be aware of the kinds of mistakes that we have identified and become more vigilant. OCLC members who are cataloging new records might consider the option of transferring new records to an ASCII file through the Savescreen option, calling them up in a word processing program with a spell checker, and eliminating these problems ahead of time.

Now that the technology has arrived for keyword searching of bibliographic utilities, those institutions should delete misspellings from their records as part of their ongoing maintenance. Perhaps it might be possible for them to run new records through a spell checker on a daily basis.

Finally, we would like to see a follow-up study performed on a larger database to verify our findings about predicting the types of words that are prone to misspellings.

REFERENCES

1. Thomas A. Peters, "When Smart People Fail: An Analysis of the Transaction Log of an Online Public Access Catalog," *Journal of Academic Librarianship*, 15:267–73 (Nov. 1989).
2. Ron Blazek, "Problems with OPAC: A Case Study in an Academic Library," *RQ* 28:169–78 (Winter 1988).
3. "The Dirty Database Test," *American Libraries* 22:197 (Mar. 1991).
4. Charles P. Bourne, "Frequency and Impact of Spelling Errors in Bibliographic Data Bases," *Information Processing and Management* 13:1–12 (1977).
5. Cynthia C. Ryans, "A Study of Errors Found in Non-MARC Cataloging in a Machine-Assisted System," *Journal of Library Automation* 11:125–32 (June 1978).
6. Joseph J. Pollock and Antonio Zamora, "Automatic Spelling Error Detection and Correction in Textual Databases," *Proceedings of the ASIS Annual Meeting* 19:236–38 (1982). J. J. Pollock and A. Zamora, "Collection and Characterization of Spelling Errors in Scientific and Scholarly Text," *Journal of the American Society for Information Science* 34:51–58 (Jan. 1983). J. J. Pollack and A. Zamora, "System Design for Detection and Correction of Spelling Errors in Scientific and Scholarly Text," *Journal of the American Society for Information Science* 35:104–9 (Mar. 1984). A. Zamora, "Control of Spelling Errors in a Large Database," *Proceedings of the ASIS Annual Meeting* 15:13–17 (Nov. 1978). Antonio Zamora, "Automatic Detection and Correction of Spelling Errors in a Large Data Base," *Journal of the American Society for Information Management* 31:51–57 (Jan. 1980). E. M. Zamora, J. J. Pollock, and A. Zamora, "The Use of Trigram Analysis for Spelling Error Detection," *Information Processing and Management* 17:305–16 (1981).
7. E. T. O'Neill and D. Vizine-Goetz, "The Impact of Spelling Errors on Databases and Indexes," *Proceedings of the National Online Meeting 1989* (May 1989), p.9–11. ■■

Special Section: Happy Birthday to MELVYL® (Part 1)

The MELVYL System: The Next Five Years and Beyond

Michael G. Berger

Planning for the next five years of the MELVYL system is described in the context of University of California information system planning. The planning environment is outlined from which are derived the objectives for the continued growth of the MELVYL system. The technical evolution of the MELVYL system necessary to meet the objectives is also reviewed. Envisioned in this technical evolution is the conversion of the MELVYL system to a client/server architecture that includes a graphical interface. Future plans for the MELVYL system provide a basis for tackling the problems of fragmented databases and information overload. Four initiatives to alleviate these problems are briefly described.

Over the last ten years, the MELVYL system has evolved into a key element in the university's plan for access to information. The University of California's strategy for library automation stresses the need for universitywide access through the continued deployment of University of California networks, the development of campus-integrated library automation systems, and the expansion of the MELVYL system into an information utility. The present strategy reaffirms the "one University, one library" objective of the *Plan for Development, 1978–1988*, the ten-year plan that led to the development of the MELVYL catalog.[1]

The location of information resources, once constrained by technical limitations, can now be determined by the needs of users, the quality of services, the economics of procuring the resources, and the strength of networks. The MELVYL system is only one component of the university's effort to coordinate access to scholarly information, a role it shares with the other components—campus libraries, computer centers, and departments. Future growth depends on coordination and cooperation among all components.

This paper describes options for furthering the role of the MELVYL system in a continually evolving environment. First, the current environment and planning assumptions about the future are reviewed, placing the growth of the MELVYL system in the context of the overall planning for the development of university automated information systems. Next, the objectives underlying the continued development of the MELVYL system are proposed. These objectives suggest a shift of emphasis by giving the MELVYL system a more active role in mounting databases, making more effective use of network resources, and incorporating new technologies for access and the display of information. Then, the technical evolution of the MELVYL system necessary to achieve its role in the University's information structure is described. Finally, challenges beyond the next five years are suggested.

CURRENT ENVIRONMENT AND PLANNING ASSUMPTIONS

At the University of California, users access information resources through a mosaic of loosely connected systems. The university's strategy for access to information differentiates where information is stored from where it is accessed. Though the physical storage of information may be centralized within the MELVYL system, at the campus, or at the national level, access should be universal, lim-

Michael G. Berger is Assistant Director for Planning, Division of Library Automation, University of California, Office of the President, Oakland, California.

ited only by user affiliation as appropriate for some restricted data such as personal files.

Within this framework, the MELVYL system will continue to provide access to the monographs and periodicals holdings of the university and a broad spectrum of abstracting and indexing (A&I) databases covering the major disciplines of interest to the university community. In addition, access may also be provided to nonbibliographic information sources, such as source material in electronic form, electronic journals, and selected scientific and image databases of general universitywide interest.

The strategy for coordinated access to information resources maximizes the university's investment in existing systems by drawing together the various aspects of the present environment into a coordinated system. The major components of the present environment that are influencing planning include the following.

THE MELVYL SYSTEM

With the addition of A&I databases, the MELVYL catalog has made a transition from a catalog to a system. In addition to requiring access to a varied list of information sources, users are also requesting that the MELVYL system connect to network-based services, such as electronic mail and remote printing. Investments in these network infrastructure services will then facilitate development of links to document delivery systems and current-awareness functions. Another aspect of the MELVYL system's role as an information utility is demonstrated by the growing use of MELVYL services by non-UC institutions such as Stanford University, the California Academy of Sciences, and the California State Library. In this regard, the University of California has established reciprocal arrangements with Stanford to trade access to proprietary databases. Together, UC and Stanford have negotiated contracts with database providers that allow joint access.

Use of the MELVYL system continues to expand (see figures 1–5). Though there has been a drop-off in the use of the MELVYL monographs database as campuses have installed local library systems that include online catalogs, this has been more than offset by the use of the MELVYL journal articles databases. In addition, remote use of the catalog has grown dramatically. The monograph

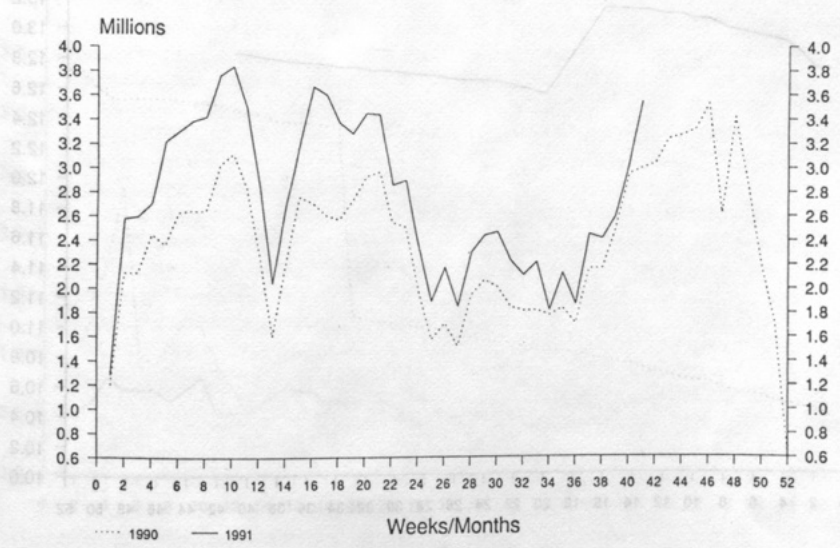


Figure 1. Number of Records Displayed.

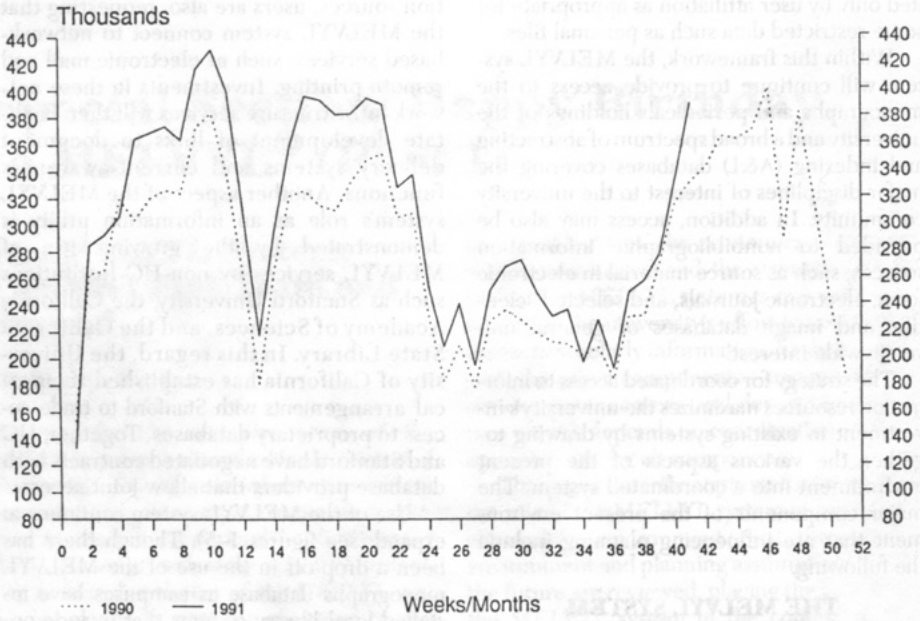
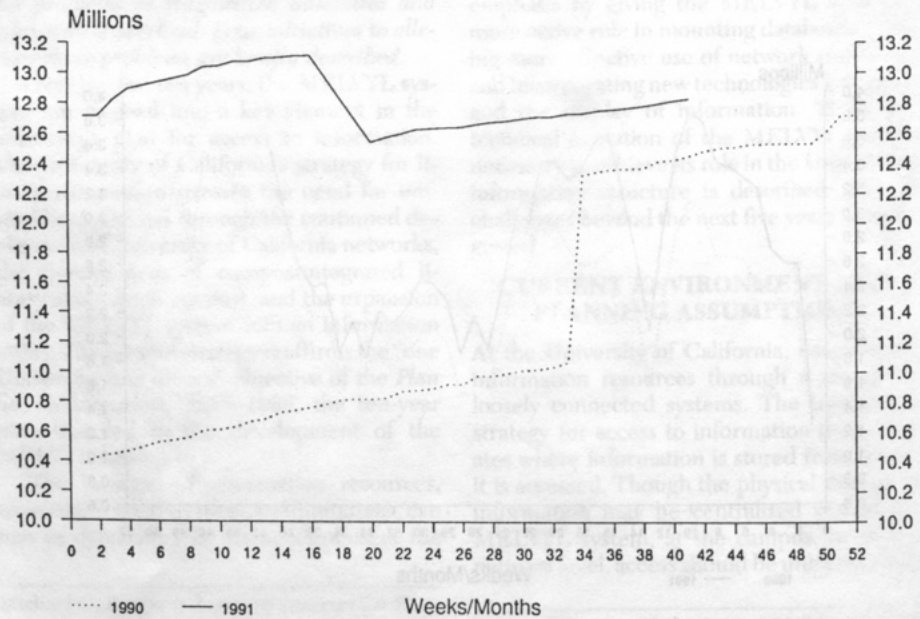


Figure 2. Number of FIND Commands Issued.



(The apparent dramatic increase in September 1990 and the apparent dip in May 1991 are both due to changes in the algorithm used to estimate holdings; loading has actually held fairly steady.)

Figure 3. Database Size—Holdings.

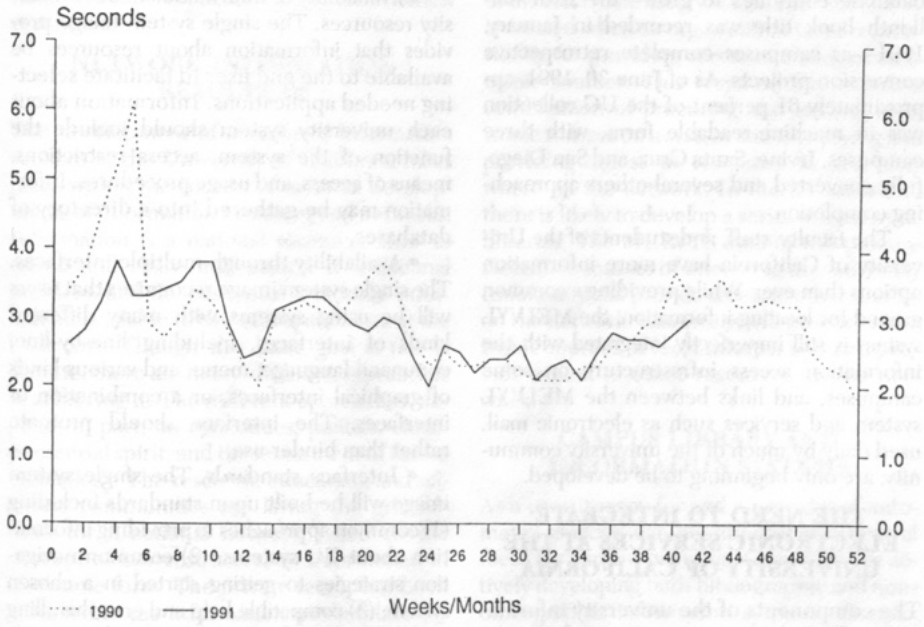


Figure 4. Search Response Time.

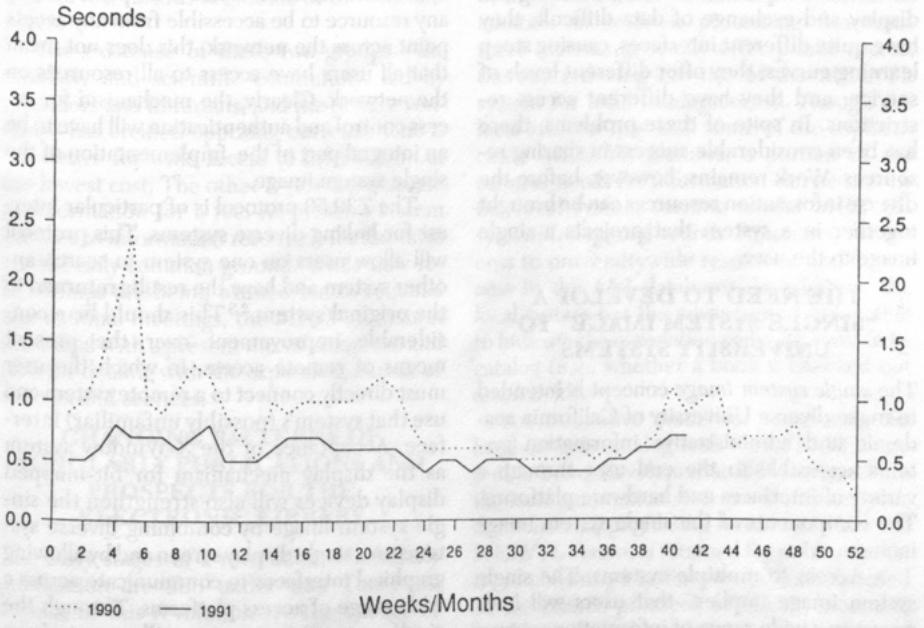


Figure 5. Display Response Time.

database continues to grow—the sixth millionth book title was recorded in January, 1991—as campuses complete retrospective conversion projects. As of June 30, 1991, approximately 81 percent of the UC collection was in machine-readable form, with three campuses, Irvine, Santa Cruz, and San Diego, fully converted and several others approaching completion.

The faculty, staff, and students of the University of California have more information options than ever. While providing a common ground for locating information, the MELVYL system is still imperfectly integrated with the information access infrastructure on some campuses, and links between the MELVYL system and services such as electronic mail, used daily by much of the university community, are only beginning to be developed.

THE NEED TO INTEGRATE ELECTRONIC SERVICES AT THE UNIVERSITY OF CALIFORNIA

The components of the university information infrastructure—computers, information sources, networks, and electronic services—are slowly becoming integrated into a common system. Existing systems, however, do not share compatible standards, making the display and exchange of data difficult; they have quite different interfaces, causing steep learning curves; they offer different levels of service; and they have different access restrictions. In spite of these problems, there has been considerable success in sharing resources. Work remains, however, before the diverse information resources can be brought together in a system that projects a single image to the user.

THE NEED TO DEVELOP A "SINGLE SYSTEM IMAGE" TO UNIVERSITY SYSTEMS

The *single system image* concept is intended to make diverse University of California academic and administrative information systems accessible to the end user through a variety of interfaces and hardware platforms. The components of the single system image include:

- Access to multiple systems. The single system image implies that users will have access to a wide range of information sources and services under university control, as well as to regional and national networks.
- Availability of information about university resources. The single system image provides that information about resources be available to the end user to facilitate selecting needed applications. Information about each university system should include the function of the system, access restrictions, means of access, and usage procedures. Information may be gathered into a directory of databases.
- Availability through multiple interfaces. The single system image recognizes that users will be using systems with many different kinds of interfaces, including line-by-line, command language, menu, and various kinds of graphical interfaces, or a combination of interfaces. The interface should promote rather than hinder use.
- Interface standards. The single system image will be built upon standards including (1) common approaches to providing information about UC systems; (2) common navigation strategies to getting started in a chosen system; (3) compatible help and error-handling strategies; (4) information retrieval protocols such as Z39.50; (5) display protocols, such as the X-Window system; and (6) compatible access to services such as printing.
- Access control. While UC's plans call for any resource to be accessible from any access point across the network, this does not mean that all users have access to all resources on the network. Clearly, the mechanism for access control and authentication will have to be an integral part of the implementation of the single system image.

The Z39.50 protocol is of particular interest for linking diverse systems. This protocol will allow users on one system to search another system and have the results returned to the original system.[2,3] This should be a considerable improvement over the present means of remote access, in which the user must directly connect to a remote system and use that system's (possibly unfamiliar) interface. Acceptance of the X-Window system as the display mechanism for bit-mapped display devices will also strengthen the single system image by combining diverse systems on a single display screen and by allowing graphical interfaces to communicate across a wide range of access platforms. Through the single system image, users will come to have stable expectations about university systems and be able to rely on these expectations to

deal with new applications and information resources.

NETWORK INFORMATION RESOURCES

The development of automated information services at the University of California takes place within the context of national developments of information resources and networks. Information is a national resource. How to create and share information is a national problem, and the basic goal is to make information available to the widest audience at the least cost. Though this basic goal is not in dispute, there are many different approaches to achieving it. The different approaches depend, in part, on economics, politics, entrepreneurial spirit, and the assessment of future technology. The research, education, and scientific communities are laying the groundwork for a national network based on the concept of low-cost access to information in electronic form. The private sector has been attracted to the information marketplace for profit because of the high value organizations and individuals place on locating and using information. Both as database creators and providers, the private sector has targeted and served various segments of the marketplace.

The existence of these two groups—the scholarly and scientific community, led by educators and librarians, and the private vendors—has created opposing currents. One is the desire for wide access to information at the lowest cost. The other is restricted access to information for a fee. At present, system licenses to information resources for flat rates are the only common ground. While new relationships are being worked out in regional and national meetings, the MELVYL system must deal with a present mix of resources and plan for dealing with the expanding mosaic of sources and services.

CHANGING PATTERNS OF SCHOLARLY PUBLICATIONS AND PROSPECTS FOR THE ELECTRONIC LIBRARY

The early stages of a revolution in scholarly publication are also under way. The main themes of this revolution are the storage of source materials in electronic form and the development of electronic journals. The result of this revolution will be the development of the electronic library. The electronic library is created when an organization takes responsibility for providing access to a body of electronic sources. This responsibility entails a combination of procuring and loading electronic sources on a system and identifying and providing access to resources loaded elsewhere on the network. In the UC situation, there is likely to develop a series of electronic libraries. The MELVYL system will procure electronic sources of interest to the university community and provide access to other sources on the network. Campuses will procure electronic sources of local interest and also provide access to other resources on the network.

CAMPUS LIBRARY AND INFORMATION SYSTEMS

As a constituency for and a provider of automated services, campus libraries are a central focus for current planning. Campuses are actively developing both bibliographic and nonbibliographic information systems. Campus libraries are achieving significant improvements in service, internal efficiencies, and reductions in the growth rate of library processing costs through the implementation of integrated library automation systems. In addition to internal processing functions, these systems generally offer public access, thereby reducing the use of the MELVYL system for local catalog access. Though the local systems may lack some of the sophisticated MELVYL features, a portion of the campus needs for information can be met by the local system turning to the MELVYL system for special search functions, for access to universitywide resources, and for access to the A&I databases. In addition, the local system has the advantage of being able to indicate the circulation status of items in its catalog (e.g., whether a book is checked out or on hold).

An alternative model for integrating the local campus catalog with the MELVYL system is being developed at UC Davis, where the local automation system will primarily address internal processing functions, while the MELVYL system is retained for public access. By establishing a system-to-system link based on the Z39.50 protocol, MELVYL users will be able to determine the circulation status of Davis items from the local circulation control system.

CHANGING USER EXPECTATIONS

The beneficiary of UC-wide and national planning initiatives is the user and, in particular, the end user. Information retrieval was once viewed as the purview of the librarians or information specialists who mediated between the end user and the information source. The advent of the online catalog and the easy availability of online resources has shifted the emphasis to the end user. The online catalog has also led to increased expectations and demands for more resources and services.

PLANNING OBJECTIVES FOR MELVYL SYSTEM DEVELOPMENT

MELVYL system objectives are proposed that extend its current status as an information utility combining database access and electronic services through a common interface. These objectives recognize that while there may be a common interface, the resources may be housed on the MELVYL system, local campus systems, or anywhere on the network. Specific objectives are to:

- Continue growth of the catalog and periodical databases. The catalog and periodicals database will continue to grow with current acquisitions and retrospective conversion. Approximately 81 percent of the university's collection have been converted.
- Accelerate loading of A&I databases. The loading of A&I databases will be accelerated, with the goal of loading a core of twelve to fifteen databases supporting the major disciplines in the next five years. UC will maintain the flexibility to respond to changes in the composition of A&I databases by being able to add or cancel licenses as needed to maintain the best and most cost-effective available resources for UC needs.
- Provide access to specialized databases. A considerable number of specialized databases are being created at the university and other institutions. Although there is no current project to mount specialized databases, the Division of Library Automation will work with database producers to provide access to these resources.
- Provide access to electronic source databases. Providing access to full-text databases of source materials is the logical extension of the MELVYL system's role as an access system. Presently, planning is under way to load the full text of a portion of the Information Access Corporation's databases now indexed on the MELVYL system. An immediate priority will be for the MELVYL system to provide access to page image files of electronic source information for transmission across the network for high-quality access and printing.
- Provide access to electronic journals. Electronic journals are a special case of source materials, as there may be no printed edition. Electronic journals will be mounted on the MELVYL system consistent with the university's acquisition policies. The MELVYL system will also be able to provide access to electronic journals mounted at other locations on the network as these journals become available. As an electronic library, the MELVYL system may act as an archive for issues of electronic journals.

PROVIDE ACCESS TO REFERENCE SOURCES

Considerable reference material is becoming available in electronic form, such as dictionaries, encyclopedias, handbooks, and pharmacopoeias. The MELVYL system is a candidate for providing access to a subset of these resources of interest to the academic community.

- Provide access to a directory of databases. The number of database sources continues to grow at an explosive rate. Except with the most common databases, users have difficulty determining what databases exist as well as their coverage, availability, access restrictions, and methods of use. Presently users accessing the MELVYL system are given a choice of databases loaded centrally and a selection of databases available on the network. At some point, there will be a critical mass of these databases available through the network, making it necessary to localize information about these resources in a directory of databases that users will be able to search by provider, title, and topic.[4,5]
- Plan for intersystem standards. The university is committed to implementing standards that will facilitate access to and sharing of data between systems. There is already wide-scale acceptance of standards, such as MARC and TCP/IP, with OSI waiting in the wings. Soon, the computer-to-computer information retrieval protocol Z39.50 will be implemented to link diverse information services. Other standards for data interchange,

authentication, and printing are in the process of being discussed, developed, and implemented.

• Facilitate data coordination and data quality. The records in the MELVYL union catalog and CALLS (California Academic Library List of Serials) are the result of the work of campus libraries. The MELVYL system will continue to protect the integrity of these records and coordinate the MELVYL catalog databases with local campus files as well as to upgrade them and to extend the value of the UC bibliographic file by incorporating information from national bibliographic sources, such as the LC MARC and CONSER files.

• Plan for delivery of electronic services to the UC community. Planning is under way to develop an operational framework to provide electronic services in the UC network environment. These services will include electronic mail for transmission of requests for services such as document delivery, various kinds of file transfer mechanisms, and remote printing.

• Plan for conversion of the MELVYL system for use in a workstation environment. Creating a workstation environment implies a conceptual, software, and hardware shift from terminals to intelligent front ends that integrate internal processing capabilities with standardized interfaces such as X-Windows, Z39.50, and file transfer protocols.[6–8] In this environment, access to information resources is merged with the other processing functions of the faculty, staff, and student workplace. It is important to recognize that *workstation*, in this context, is a general term, encompassing today's UNIX workstations (from vendors such as DEC, IBM, SUN, NEXT, and Hewlett-Packard), PCs, Macintoshes, and others to come. Any strategy for workstation support must recognize the present and future diversity of the installed base of such workstations. An interesting example of a workstation used as an intelligent front end to the MELVYL system is provided by Michael Buckland's OASIS project, which does more refined processing of records retrieved and downloaded from the MELVYL system.[9]

• Maintain adequate MELVYL performance. During the transition from access to new information sources and new access technologies, the MELVYL system must maintain adequate performance in terms of reliability, response time, and usability.

• Continue planning for disaster recovery and system backup. Planning for disaster recovery and system backup has been made more complex by the multiplication of MELVYL databases, the implementation of campus catalogs, the development of alliances between organizations whose systems may be able to provide backup during disasters, and by the development of technical standards such as Z39.50 that will facilitate resource sharing. During the next five years, work will continue on developing plans and policies for disaster recovery that reflect the mosaic of databases to be backed up, identifying sites that could be used in times of emergencies, and establishing agreements with other database providers for reciprocal emergency access.

Planning for the future growth of the MELVYL system occurs in the context of overall planning for libraries and other automated information activities. Plans are reviewed widely by university administrative groups, librarians, library staff, and faculty. Decisions such as which databases to mount on the MELVYL system, changes to the interface, and new functions are made in consultation with standing and ad hoc committees.

TECHNICAL EVOLUTION OF THE MELVYL SYSTEM

During the next five years, the MELVYL system will evolve into a network information server based on national and international standards. In the network information server environment, the database and retrieval software will be separated from the user interface with the Z39.50 protocol used to communicate between the database and the interface. A client/server architecture will allow the system to incorporate new technologies more flexibly; the result will be a set of services that is far more agile than today's configuration when faced with new service demands, new technologies, and shifting cost-technology tradeoffs. The approach will be evolutionary, adding new capabilities such as workstation support, while retaining existing services.

One possibility will be to employ multiple database management systems, selecting one specialized for bibliographic databases, another for full-text databases, a third for image databases, and a fourth for scientific/numeric data to obtain maximum leverage through the use of commercially available software. There

is a very high cost to such diversity, not just in terms of licensing and maintaining software but also in terms of staff and training. A rigorous technical and cost-benefit evaluation of database management technology available on the marketplace in the 1992–93 time frame will be required to make these decisions. A client/server environment requires a revision of the MELVYL user interface.

The present MELVYL interface is really a composite of interfaces that has evolved as the MELVYL system has added new functions and databases. To take advantage of new technologies and facilitate the future growth of the MELVYL system, a graphical user interface is proposed to consolidate the differing approaches of the present interface into a single, unified interface. Graphical interfaces facilitate learning and using online catalogs and improving the presentation of existing functions, as well as providing a platform for new approaches to manipulating information. A graphical interface offers the user opportunities for adapting the interface to personal style, allowing both graphical and command line approaches.

As part of the development of the graphical interface, the present line-by-line interface will be revised as well. A single interface will be conceptualized with a graphical and line-by-line component. A number of factors warrant the continuation of a line-by-line interface. First, during the transition period to graphical environments, the availability of a line-by-line interface will provide continuity for current users. Second, some catalog system functions are better handled in a command line environment. Third, there will be a continuing need for a line-by-line interface as users will be likely to access the system from a combination of graphical and ASCII terminal locations, such as mixing graphical access at work with ASCII access from home. Fourth, there is likely to be a lengthy period of transition to graphical platforms in UC libraries. Revising the current interface to be compatible with the graphical interface will give all users immediate benefits from the proposed revisions to the interface.

Graphical interface is a loose term applied to windowed interfaces typically activated by a mouse and keyboard and cued by graphical devices such as buttons and icons. The designation *graphical interface* has been applied to simple systems with little more than a window capability as well as to complex systems with multiple layers of windows and arrays of buttons and icons. Proposed here is a graphical interface utilizing the X-Windows technology. It is reasonable to assume that systems based on X-Windows technology will predominate in the academic community. Though presently X-Windows capabilities imply expensive workstations, the X-Windows protocol is independent of hardware and can be utilized in specialized terminals (X-terminals) connected to networks and by PCs configured as X-terminals.

The complexities of new access and service environments require improved interfaces to keep up with the evolution of existing systems and the promise of new ones. Conceptually, an interface should be more than its display mechanisms. The design of the proposed interface will be based on the needs and capabilities of its users.

THE MELVYL USER INTERFACE: DESIGN PRINCIPLES

To provide some stability in the rapidly changing user interface environment, a set of design principles are proposed to guide and evaluate interface approaches. These include the following:

- The interface must be comprehensive. The interface must provide coordinated access to all MELVYL system functions including those that link the MELVYL system to others across networks.
- The interface must provide an orderly transition between the novice and expert user. The interface assumes that users will become more expert with the MELVYL system through use, and learning paths toward more expert use should be provided. The terms *novice* and *expert,* in the context of the interface, apply to the overall expertise of the user and not just to the selection of a menu mode or command mode for communicating with the system. Thus, whether commands or menus are used is less important than the functions attempted.
- The interface must provide a range of communication approaches. In the proposed interface, the user must have a range of communication approaches, from mouse-activated graphical user interfaces to command-oriented interfaces.
- The user must be able to configure a number of elements in the interface. As pro-

vided by X-Windows and the MELVYL system application, users should be able to modify the appearance of the interface and the location of windows, icons, and buttons. In the present MELVYL system, users have the option of setting a number of parameters for each session, such as their database choice, the date range and language of their search, and the default display. In the graphical interface, these options could be cued graphically.

• The user must be able to make informed decisions about system use. The proposed interface should facilitate user decision making by providing information about the user's session. Navigational paths should be functionally organized, expanding from the simple to the complex.

• Actions in the interface must be distinctive and memorable. The placement of windows, use of graphics and color, the location of text, and the use of different typefaces and other devices will be used to reinforce the user's sense of the functions of the system and his or her current location in the system.

• The interface should expand the user's working memory. By placement of information on the screen, the proposed interface can facilitate the use of MELVYL functions. Because recognition memory is robust and active, having graphical and text reminders should aid the user's navigation of the interface.

• The interface must facilitate use of the MELVYL system without being obtrusive. While graphical aids, windows, and icons can facilitate system use and future learning, these communication techniques should not dominate the interface or become so idiosyncratic that it becomes difficult for the user to transfer knowledge about use of the catalog to other systems with similar functions.

• The interface must be multitasking. The user should be able to do more than one task at a time, such as downloading a bibliography while doing another search.

These general principles are the starting point in evaluating existing interfaces and will provide a framework for development of the MELVYL graphical interface.

The graphical user interface is more than just an improvement in display technology; it is a change in the way the interface is viewed by both designers and users. The interface goes beyond sequential processing of a single application or application task to become an asynchronous window on multiple applications. Users can move from one application to another in a manner based on their own work habits rather than by system-determined sequences. By jumping from window to window, users can run multiple tasks simultaneously and can combine elements of different tasks into a single task. What makes this flexibility possible is the independence of the interface from the application and the inherent capabilities built into the graphical interface.

THE MELVYL SYSTEM: CHALLENGES AT THE PLANNING HORIZON

The MELVYL system is very much a creature of its environment. It has been developed to meet the needs of its community, and the system has evolved as the environment has changed. What was once a replacement for the card catalog has become an information utility providing access not only to the traditional library catalog but also to journal literature and other network information sources. The development of a graphical interface will allow the MELVYL system to stay current with the best of advancing technology. In the next five years, the MELVYL system will become more of a component in the university's coordinated access to information than a stand-alone system.

The next five years will also be a transition period, bringing the system to the point where solutions can be sought for persistent challenges to effective information access. These challenges involve changing the way information is viewed and the role of the catalog in the information-seeking process. Presently, access to information is fragmented by database, type of publication, type of index, date of publication, and by the user's location. In one database, users are finding more information than they are willing to display. In the MELVYL system, users retrieve an average of one hundred records per search in the catalog database but only display about fifteen records per search. Because the most commonly used display command lists retrievals in alphabetical order by author and title, users' displays are heavily skewed toward author names beginning with the first letters of the alphabet. If users are overloaded with information from a single database, they are truly overwhelmed from the cumulative resources across databases.

A major challenge for the future is to increase the scope of information searched across databases but at the same time make it possible for the user to display records that best meet the particular information need. Some of possibilities to meet this include:

- Widening the scope of the search. A user's search should be processed against all reasonable information sources. Users are often not clear on the distinctions that divide databases or even indexes within a database. For example, users search for periodicals and journal articles in the monograph database, books in journal article databases, and corporate authors as personal authors or titles. Users exhaust themselves searching for a topic in one database while another database may have more relevant materials.
- Providing more tools for the user to focus the search. Even in the single database, the user is often swamped with information. Searches against multiple files will compound the problem unless better tools are provided, such as entry vocabularies for formulating searches, a feedback mechanism for redirecting searching, and limiting functions for reducing large retrievals.
- Providing more meaningful information about retrieved publications. There is considerable overlap between databases. For example, an article in one database may only have the citation, in another the author's address, in another the abstract, in another a review of the article, and in another the full text. Links need to be established between like works in different databases so that users will be able to retrieve the full extent of the available information.
- Aiding the user in recognizing the most significant literature. Presently, the results of all searches are treated equally. Not all works are created equal; some are acknowledged as classics in their fields. Means need to be developed to permit users to determine the standing of retrievals. Users need to be able to determine the core or classic works in a field, to consult reviews of works, and to learn about the authors of works by links to biographical information. Much of the information needed to make these kinds of judgments is available now in machine-readable form and must be gathered into more holistic systems.

Work on these challenges will provide opportunities for libraries and librarians. The online catalog has gone a long way in achieving the first two objectives of the catalog first proposed by C. A. Cutter: (1) to enable a person to find a book by author, title, or subject; and (2) to show what the library has by author, subject, or by kind of literature.

The future, however, may be in meeting Cutter's third objective: "To assist in the choice of work . . . as to its edition (bibliographically) . . . [and] . . . as to its character (literary or topical)."[10] A reorientation of the online catalog can assist the user by breaking down the barriers to effective searching caused by the online catalog itself, by the unfamiliar details of bibliographic representation of works, and by the often rigid assignment of subject headings. Where the online catalog cannot overcome these barriers by allowing simple searching of multiple databases or multiple indexes within a database, the catalog should reveal the structure of the system and the underlying data so that users can learn to make better searching decisions. Finally, the online catalog can provide a learning environment for the user to become self-sufficient.

ACKNOWLEDGMENTS

This article has been drawn from "The Role of the MELVYL System in the University of California Library Automation: The Next Five Years,"[11] a planning document prepared for the University of California under the direction of Clifford Lynch, the director of Library Automation, and the "Development of a Prototype Graphical Interface for the MELVYL System," prepared by Michael Berger.[12] These documents were extensively reviewed by Division of Library Automation and other University of California staff, who contributed in substance and detail.

REFERENCES

1. *Universitywide Library Planning. The University of California Libraries: A Plan for Development 1978–1988* (Berkeley, Calif.: Office of the Executive Director of Universitywide Library Planning, Systemwide Administration, University of California, 1977).
2. Michael K. Buckland and Clifford A. Lynch, "The Linked Systems Protocol and the Future of Bibliographic Networks and Systems," *Information Technology and Libraries* 6, no.2:83–88 (June 1987).
3. Michael K. Buckland and Clifford A. Lynch,

"National and International Implications of the Linked Systems Protocol for Online Bibliographic Systems," *Cataloging & Classification Quarterly* 8, nos.3/4:15–33 (Spring 1988).
4. Mary E. Engle and Clifford A. Lynch, *Directory of Electronic Information Resources: A Feasibility Study*, Technical Report #4 (Oakland, Calif.: Division of Library Automation, Information Systems and Administrative Services, University of California Office of the President, 1990).
5. Clifford A. Lynch and Cecilia M. Preston, "Describing and Classifying Networked Information Resources," *Electronic Networking* 2, no.1:10–12 (Spring 1992).
6. Clifford A. Lynch, "The Client/Server Model in Information Retrieval," in *Interfaces for Information Retrieval and Online Systems: The State of the Art*, ed. Martin Dillon (Westport, Conn.: Greenwood, 1991), p.301–18.
7. Clifford A. Lynch, "Online Searching on the Internet: The Challenge of Information Semantics for Networked Information," in *Proceedings of the National Online Meeting, New York, N.Y., May 5–7, 1992* (Medford, N.J.: Learned Information, 1992).
8. Clifford A. Lynch, "Networked Information: A Revolution in Progress," in *Proceedings of the Lancaster Clinic: Networks, Open Access, and Virtual Libraries: Implications for the Research Library* (Urbana, Ill.: Graduate School of Library and Information Science, University of Illinois, in press).
9. Michael K. Buckland, Barbara Norgard, and Christian Plaunt, "Design for an Adaptive Library Catalog," in *Proceedings of the ASIS Mid-Year Meeting, Albuquerque, NM, May 27–30, 1992* (in press).
10. C. A. Cutter, reprinted in *Rules for a Dictionary Catalog*, 4th ed. (Washington, D.C.: Govt. Print. Off., 1904), p.12.
11. Division of Library Automation, *The Role of the MELVYL System in University of California Library Automation* (Oakland, Calif.: Division of Library Automation, Information Systems and Administrative Services, University of California Office of the President, 1991).
12. Michael G. Berger, *Development of a Prototype Graphical Interface for the MELVYL System* (Oakland, Calif.: Division of Library Automation, Information Systems and Administrative Services, University of California Office of the President, 1992). ■■

Agenda for Online Catalog Designers

Michael K. Buckland

Fifteen recommendations are offered for the improvement of online catalogs within the categories of closer connections to the users' work environment, SDI, downloading, reform of LCSH, *enhanced search capabilities, and linking with other bibliographies and text.*

Recognition of the achievements of the first ten years of the MELVYL online system of the University of California occasions an excellent opportunity to examine what needs to be done in the next ten years of online catalog design and development. What follows is a personal selection of improvements not only for the MELVYL system but for online catalogs generally.

Michael K. Buckland is Professor of Library and Information Studies, University of California, Berkeley. During 1983–86, as Assistant Vice-President for Library Plans and Policies in the University of California systemwide administration, his responsibilities included the Division of Library Automation and the MELVYL system.

USER ENVIRONMENT

The online catalog has two quite different kinds of impact. For all who visit the library, it is a different sort of catalog, with a keyboard, screen, and a new way of searching that replaces passive trays of cards.

A different impact arises with the growing proportion of library users whose work habits and working environments have changed to include routine use of computers. For these persons, the option of remote access to the library's catalog has constituted an important new extension of library service. Not since library catalogs were (infrequently) printed and distributed in book form in the nineteenth century has this kind of catalog access been possible. This second impact is selective, an enhancement of service for those whose work habits and equipment enable them to benefit. Library automation to improve library service within the library is clearly

useful. However, the ability of the library to arrange for access from outside the library to materials stored electronically, such that users with suitable equipment and skills can use these resources by themselves, constitutes a much more substantial extension of library service.

Because people have moved to a personal computing environment for their work, they need the provision of online access to the online catalog, online bibliographies, and any other online resources because the effective performance of their work is based on access to electronic records. Their work is constrained if such access is not provided. For this reason library automation, hitherto based on factors internal to the library, should now be associated with and paced by the parallel shift in the "task environment" of the people the library serves. Once library users begin to work electronically, they are hindered by the lack of remote access to an online catalog and to materials in electronic form. This close coupling of library development with changes in users' working styles requires a new perspective. Any serious agenda for automation in library service should include enhancements designed to bring service to where the users are and into their personal working (and computing) environment. Our first four agenda items are in this class.

Automatic SDI

The Selective Dissemination of Information (SDI) is the notification of library users of selected, newly received items relevant to their personal interests. SDI is a well-established practice in small, specialized libraries but is labor intensive and, therefore, rarely found in large libraries. The idea of SDI has found new currency outside of libraries as "information filtering." The (largely independent) developments of electronic mail and of online library catalogs can be combined to provide automatic SDI if the catalog has an "AND LOADED SINCE [date]" search limit capability (as the MELVYL system does) or can achieve a similar effect through, for example, record ID numbers in consecutive order.

One feasible approach would be along the following lines. A library user's SDI profile can be expressed in terms of an online search statement (e.g., FIND SUBJECT CATALOGS, ONLINE) and identified by the user's electronic mail address (e.g., buckland@otlet.berkeley.edu). During off-peak periods, at intervals such as once a month, an SDI program would initiate each search with the AND LOADED SINCE search limit set so as to capture records added to the catalog since the previous running of the program. Each search result would then be sent automatically as electronic mail to that library user's e-mail address. Implementing such a service would build on disparate, existing investments in e-mail systems, telecommunications networks, and an online catalog and would add to the value received from each.

It was clear by the mid-1980s that such a service would be feasible and relatively simple to implement, but it was not perceived as a priority for the MELVYL system at that time.[1] The practice of loading indexing and abstracting files in conjunction with online catalogs increases the scope for what could be a popular and inexpensive service.

Formatted Downloading

Library users, especially users of academic libraries, have developed personal computing environments, primarily for word processing, for the very same kind of work that generates library use. Dial-in access to library catalogs has become standard, and standard communications software allows for catalog searches to be downloaded. However, identifying individual records within the downloaded stream of characters and distinguishing the author, title, and other fields within the downloaded records usually requires tedious editing. Software for this is available for some online catalogs and is under development for others, but the widespread and prolonged failure to provide users with software that retains the formatting of downloaded records reflects poorly on attitudes toward library users.

High-Speed Record Transmission

A decade ago the MELVYL system and other catalogs were designed to communicate with plain terminals over relatively slow lines. Now that situation is changing rapidly as high-speed networks and increasing numbers of workstations are coming into use. The MELVYL system, for example, currently downloads records at no more than around one record per second in off-peak periods and significantly more slowly during busy periods. Thus, for example, downloading a set of six hundred records for further analysis in a

workstation takes at least ten minutes and possibly much longer. Effective use of the Internet, workstations, and the Z39.50 Search and Retrieve protocol will depend on much higher downloading speeds.

Cordless Telecommunications

The card catalog could only be used in one place, and an enormous advantage of an online catalog is that it can be used anywhere telecommunications can reach. Very useful first steps have been to allow "dial-in" access from outside the library and to place terminals in stacks and reading rooms where they are needed, as well as in the traditional "catalog hall." The need for electrical power and telecommunications cables to reach the terminal is still an expensive constraint. Library users, however, are beginning to carry around small, portable, battery-powered notebook computers that are very convenient for in-library use. It would be an obvious amenity to enable library users with portable computers to use the online catalog without needing to connect to a telephone line. Cordless telephones or, more likely, radio transmission of data packets hold great promise. The pioneering research at the Division of Library Automation, supported by the California State Library, demonstrated the feasibility of this approach. The initial motivation had been to reduce the high cost of cabling large libraries, but the approach clearly has numerous possibilities.[2]

RESOURCES

What of the catalog as a resource? Two aspects spring to mind.

LCSH Modernized

The predominant form of subject access in U.S. library catalogs is through the *Library of Congress Subject Headings* (*LCSH*), a system in which complex topics are expressed in the form of lengthy phrases (e.g. "Marmots as carriers of disease") or else as a heading extended by a set of qualifying subdivisions (e.g., "God—Knowableness—History of doctrines—Early church, ca. 30-600—Congresses"). There is considerable scope for routine updating and systematizing *LSCH*, but there are two more basic problems. Searching long, complex subject headings ("Exact subject searches"), while relatively easy to do when scanning headings already visible on cards or printed catalogs, is doubly difficult in an online catalog: One has to guess what the heading might be; and one must avoid any keying error. In second-generation online catalogs such as the MELVYL system a sensible solution is to allow subject keyword searching, atomizing the carefully constructed precoordinate *LCSH* into component words for postcoordinate searching on these fragments. Online use of the *LCSH* system could be greatly simplified if the existing provision for subdivisions were developed and systematized as a verbal faceted classification. This is hardly an avant-garde suggestion; it is essentially what Paul Otlet and Henri La Fontaine did in 1895 to develop the Dewey Decimal Classification into the more powerful and versatile Universal Decimal Classification. A more faceted approach would, like the Anglo-American Cataloging Rules, the Dewey Decimal Classification, and the Library of Congress Classification, be respectably grounded in the nineteenth century.

Cataloging Quality and Completeness

The MELVYL system is a union catalog. As a policy decision, it was decided to retain every variant detail of cataloging from several different cataloging departments. This complexity is largely hidden from the users, since only one form of the record is ordinarily displayed, but it is a wonderful boon for those who teach cataloging. Almost any MELVYL MARC display for an item cataloged by multiple catalog departments can provide a basis for discussion of record differences. There are two aspects to this: consistency and completeness. University of California cataloging may well be of above-average quality, yet the number of variant forms continues to increase. FIND EXACT SUBJECT LIBRARY, for example, yields two records: one a cataloging error for LIBRARIES, the other a miskeying for LIBERTY. A program of record cleansing at OCLC, recently reported to be correcting 30,000 records a day, should be an inspiration to all catalog administrators.

A different problem is that of completeness. Initially only a few fields were searchable: author, title, and subject headings. As online catalog software evolves, the number of different fields that can be searched increases, with the trend presumably toward being able to search all fields. However, the

usefulness of this extended functionality will be limited by the frequency with which these fields are left empty or contain what appears to be a higher cataloging error rate.

SEARCH CAPABILITY

Spelling and Plurals

A very useful feature of standard word processing software is the ability to identify spelling errors. This would be an obvious amenity for any online catalog, at least for subject searches that have retrieved no records. Further, it is pedantic and unfriendly to offer a service in which the user does not even know that success in searching may depend on using a plural (e.g., CATFISHES IN ART) or, perhaps, a singular form of a subject—or a variant spelling of a word when using the library catalog(ue)(s). That the catalog records may be inconsistent can reinforce a user's misunderstanding. For example, in the MELVYL system, FIND SUBJECT CHANSON retrieves 153 records, a result likely to be perceived as successful, but one that masks the fact that FIND SUBJECT CHANSONS retrieves 1,376 records. Only 113 records are common to these two sets.

More Searchable Indexes

Online catalogs started with the traditional access points of author, subject, and title. Since then the range of searchable access points has steadily increased to allow, for example, searching by date and language. The long-term expectation should be that eventually all data in or implicit in the MARC format will become searchable.

Cross-references

One of the major ingredients of cataloging is the systematization of syndetic structure: the web of cross-references between related terms. *LCSH* was expanded recently to provide "Use," "Use for," "See Broader term," "See Narrower term," "Related term" and "See Also." Further reform is needed in the case of tantalizing and minimally helpful guidance (e.g., "Gums and resins. See also specific gums and resins") in which the names of the breeds are not revealed.

In the case of the MELVYL system, cross-referencing has been implemented for names but not yet for subjects. As one should expect, "Mark Twain" will also retrieve "Samuel Clemens." It is nonsensical that in a catalog for the general public "Vietnam War" does not retrieve material on what the *LCSH* coyly calls the "Vietnamese Conflict." *LCSH* has a cross-reference, but unfortunately, the MELVYL system has yet to implement even "Use" cross-references.

Entry Vocabulary

The development of information retrieval in the past thirty years has been in two substantially different streams: traditional, deterministic, bibliographic approaches using human indexing, exact and Boolean searches, and retrieving sets of records (e.g., online catalogs and DIALOG); and probabilistic approaches emphasizing retrieval from full-text and ranked retrieval results (e.g., SMART). These streams have remained remarkably separate.[3] Online catalogs need to be designed for users who lack searching expertise and familiarity with the semantics of *LCSH*. A development that could be expected to transform the ease of use of online catalogs would be to combine these two approaches, using probabilistic techniques to derive from the user's vocabulary the terms in the system's vocabulary most likely to match the user's interests. A taste of the effects can be seen in the generic keyword search (FIND KW, actually a Boolean OR search combining title keyword and subject keyword searches), which has been implemented on some MELVYL databases, but not on the catalog. The CHESHIRE system and experiments on the OASIS system go one step further in displaying the most promising system subject headings in ranked order.[4] The effect would be similar to that of an up-to-date index, using contemporary language, to the *LCSH* and, preferably, to the LC Classification.

ADAPTIVENESS

Retrieval is more of a process than an event, so it is desirable that we think in terms of searching sessions rather than individual searches. In this context two developments are needed.

Retrieved Set Analysis

The MELVYL system, as is typical in online catalogs, indicates the number of items retrieved by any given search but does little more than that. An amenity being developed experimentally as part of the "prototype adap-

tive library catalog" project is the routine analysis of any retrieved set to provide the searcher with a summary analysis of composition of the set retrieved.[5] Such an analysis provides an informed basis for estimating the consequences of modifying the search and, therefore, for deciding what to do next. An expert searcher can, within limits, ascertain the breakdown of the retrieved set by, say, language or date, but doing so can be quite tedious, and it is a task that is very suitable for delegation to the computer. A simple command to analyze by, say, date, language, and holding library could generate a display of the profile of the material with which one is dealing. A variation on this theme is to have the system analyze the distribution of subject headings within the set retrieved by any search. This refinement is particularly useful when exploring some topic that is widely scattered over *LCSH*. A title keyword search on "Working women," for example, yields records with a very wide spread of different *LCSH*. Here, as in so many cases, the *LCSH* headings found are individually plausible, but no one would have the imagination to think of all, or even many, of them. An online catalog can be programmed to excerpt, rank by frequency, and display the *LCSH* (or, in principle, any other attribute in or implicit in the catalog records) in any retrieved set. Such display, which could be generated automatically and routinely, could provide the nonexpert searcher with a well-informed basis for deciding future moves in the search process, as well as be a useful convenience for even the most expert searcher. Adding, as some online catalogs can, counts of how many records for each heading would be retrieved from the entire database nicely complements the counts of the number in the set already retrieved. The latter indicates the options; the former, the consequences of moving the search to related headings. Expert systems can be expected to need the same kind of analytical capability as a basis for inferring and proposing good next steps to propose.

Strategic Commands

Problems arise when the complexity of a task exceeds the user's expertise. Various options may be possible, including educating the user to increase expertise; providing advice situationally; simplifying the system; providing an intermediary (human or artificial); and, as with automatic transmissions and automatic cameras, shifting some of the complexity into the system. Expert, effective searching of online bibliographic systems is done by implementing a search strategy composed of a series of tactical moves. In practice, however, not all searchers are expert. Weak expertise is associated with a lack of knowledge of search commands, search strategies, and the arrangement of material in the database. Weak expertise is a significant problem in the case of online library catalogs, which are used by untrained searchers. As the functionality of online catalogs increases, so their complexity increases and so, too, the amount of expertise needed to use them. However, very few of the available commands are frequently used. In particular, as files grow in size with the retrospective conversion of older records, the frequency with which excessive numbers of records are retrieved increases. Expert searchers know search tactics that can be used to reduce retrieved sets. The great majority of relatively inexpert users typically scroll through page after page of displayed records, then settle for the first few found or start over with some new search command.[6]

We use the term *strategic search command* to denote a search command that instructs the system to implement a series of tactical moves in some direction. Given the propensity of library users to limit themselves to only a few commands, it is difficult to see how else increasing complexity can be handled except by providing more versatile commands. As with the automatic transmission, it is a matter of enabling the user to delegate some of the complexity to the system and, as such, it is necessary that the user remain in control of the pace and direction. We recommend and are currently developing strategic commands of the form of FIND MORE, FIND RELATED RECORDS, FIND FEWER, and SUMMARIZE [the retrieved set].[7] What works for the nonexpert is also likely to be a convenient amenity for the expert.

OUT OF ISOLATION

The substitution of the new information technology for the old information of paper and card may very well be a sensible and beneficial course of action, but in the longer term it misses the point of technological change.

Sooner or later we need to rethink and redesign what is done so that it is not a mechanization of paper but fully exploits the capabilities of the new technology.[8] In this, the online catalog is of special interest. For example, online catalogs normally display retrieved records in the alphabetical order of main entry. Why? The first few displayed are not likely to be any more interesting than any others in the retrieved set. It is, perhaps, an unconscious carry-over from the necessity of filing, and therefore viewing, 3-by-5-inch cards in alphabetical order. Our last three agenda topics are directions in which the MELVYL system is already pointing.

Catalog and Bibliography

The mounting of a detailed bibliography, providing bibliographic access at the journal article level, when the MEDLINE file was loaded on the MELVYL system seemed radical at the time. Now, with hindsight, this move seems quite sensible, even overdue, yet it symbolizes the reversal of one hundred years of orthodoxy in library thinking: Catalogs are created in technical services departments with records derived from other libraries; bibliographies are normally created by publishers outside of libraries, made accessible through commercial firms, and searched by reference librarians in public services divisions. However, linking online bibliographies with online catalogs transforms this historic separation between bibliography and catalog. Linking bibliographies such as MEDLINE and CURRENT CONTENTS to holdings statements leads us toward a redefinition and dramatic enrichment of the library catalog. The new "catalog" becomes, in effect, the whole range of bibliographic access that can be linked to holdings records.[9]

Other Catalogs

That online catalogs around the world are becoming accessible at a distance over networks echoes the nineteenth-century practice of printing and distributing library catalogs in book form (which became a victim of the move to card catalogs). Facilitating access by "pass-through" (such as the MELVYL system's USE command) and, prospectively, using the emerging "Search and Retrieve" standards (NISO Z39.50; ISO 10162/10163) are valuable moves toward universal bibliographic control.

Catalog and Text

The pre-automation library was characterized by separations. The library and its catalog were more or less distant from the user's workplace, and the catalog was separate from the books. The online catalog can bring the catalog to the user and into the stacks. As files of documents become available in electronic form, texts (and other electronic objects) can be brought to the user. Further, the ability to bring both catalog and texts to the user will provide libraries with the option of having catalog records and their associated texts at the same time, engendering some of the advantages that make browsing in the stacks more attractive than doing so in card catalogs. These new connections are building the electronic library.

PRIORITIES

Since everything cannot be done at once, priorities become important. Current, unpublished research analyzing transaction logs reveals unexpectedly low levels of effectiveness in use of the MELVYL system (and probably of online catalogs generally). A user can easily spend half an hour not quite finding what a expert searcher would quickly find. Perhaps this is to be expected when a complex system is provided that nonexpert people have no choice but to use. The user's ineffectiveness should provide the major basis for priorities in online catalog development. For example, the unorthodox step of providing an "entry vocabulary" that converts the user's terminology into the system's language might do more good than any other reform for those who have to use the catalog.[10]

ACKNOWLEDGMENTS

This paper has benefited from discussions with Michael G. Berger and the assistance of Barbara A. Norgard, as well as the author's past involvement in the development of the MELVYL system.

REFERENCES AND NOTES

1. Michael K. Buckland, "Combining Electronic Mail with Online Retrieval in a Library Context," *Information Technology and Libraries* 6:266–71 (Dec. 1987).
2. Clifford A. Lynch and Edwin B. Brownrigg, *Packet Radio Networks: Architectures, Proto-*

cols, Technologies, and Applications (Oxford, England: Pergamon, 1987).
3. An unusual exception is the CHESHIRE system developed by Ray R. Larson, in which a SMART-based system operates on the conventional MARC records of a library catalog. Ray R. Larson, "Classification Clustering, Probabilistic Information Retrieval, and the Online Catalog," *Library Quarterly* 61:133–73 (Apr. 1991); Ray R. Larson, "Evaluation of Advanced Retrieval Techniques in an Experimental Online Catalog," *Journal of the American Society for Information Science* 43:34–53 (Jan. 1992).
4. For CHESIRE see Larson, "Classification Clustering" and Larson, "Evaluation." For OASIS, see Michael K. Buckland, Barbara A. Norgard, and Christian Plaunt, "Design for an Adaptive Library Catalog," *Proceedings of the American Society for Information Science Mid-Year Meeting, Albuquerque, NM, May 8–10, 1992* (forthcoming).
5. Buckland, Norgard, and Plaunt, "Design."
6. Stephen Walker, "Interactional Aspects of a Reference Retrieval System Using Semi-Automatic Query Expansion," *Informatics* 10:119–36 *Prospects for Intelligent Retrieval* (London: Aslib; 1990).
7. Buckland, Norgard, and Plaunt, "Design."
8. Michael K. Buckland, *Redesigning Library Services: A Manifesto* (Chicago: American Library Assn., forthcoming).
9. Michael K. Buckland, "Bibliography, Library Records, and the Redefinition of the Library Catalog," *Library Resources and Technical Services* 33, no.4:299–311 (Oct. 1988).
10. Michael G. Berger, private communication, December 1991. ■■

The Evaluation of a User-Oriented System: The MELVYL System's Many Designers

Laine Farley

As the MELVYL system has matured, its designers have employed numerous ways of discerning users evaluation of the system, ranging from online user comments to formal advisory groups. This article describes how these mechanisms have evolved and how they are used by the designers, and discusses the rising expectations of users.

As the recent winter Olympics drew to a close and the commentators reminded us that the summer Olympics were a few months away, they frequently quoted the simple but powerful motto of the games: "faster, higher, stronger." This resounding phrase seemed to echo and eventually shift to "faster, bigger, farther" as I reviewed user comments, answered phone calls, and responded to electronic mail about the MELVYL system recently. The MELVYL catalog has grown from a database of approximately 750,000 records to a system of nine databases and a gateway to almost forty other systems, and the expectations of its users have grown along with it. Each new feature or addition elicits an initial euphoria followed shortly by the "why can'ts"—why can't it be faster, bigger, and go farther. As other authors in this issue have pointed out, the system's biggest fans are usually its most demanding critics.

Laine Farley is MELVYL User Services Coordinator, Division of Library Automation, University of California Office of the President, Oakland, California.

WHAT DO USERS REALLY WANT?

As we, the designers, plan and discuss each change or new capability, we try to second guess the needs, preferences, and potential sources of confusion of the user. These discussions are often intense, sometimes even emotional, since everyone believes he or she knows what the user wants. After all, everyone is a user and therefore an expert. Ultimately, each controversial topic is pushed, shaped, wheedled, tweaked, tuned, and adjusted into an amalgam of the best ideas or at least the best compromises. These diverse viewpoints are a necessary part of the process, reflecting in part the diversity of the real user population. The original design team recognized the need for a broad view when they stated their first guiding principle for the development of

the catalog: "The catalog must accommodate the entire user population ranging from experienced searchers to first-time library patrons."[1]

The designers went on to state specific guidelines for designing the MELVYL system's user-friendly patron interface.[2] These guidelines are still being followed and are the foundation for new guidelines to expand the user interface to take advantage of graphical capabilities in the future.

The real definition of what is user-friendly comes not from a set of guidelines, but from the users themselves, of course. Again, the early designers anticipated the need to record and evaluate not only what users say but what they do. They employed the one-way mirror of the behavioral scientist in the form of transaction logs to observe users actions. Users remain anonymous in the logs, but identification numbers are assigned to terminal locations so that the sequential activity of a user's session can be recorded. The logs were part of the initial evaluation of the system that was funded by the Council on Library Resources.[3] The logs also have been used to answer specific questions, such as whether users were searching for corporate authors when employing the original Author index (which combined personal and corporate author indexes), as well as to capture entire sessions for analysis of unexpected difficulties.

Two additional tools that were part of the original plan are more similar to those used by market researchers. Online questionnaires, akin to market surveys, were used to evaluate user reactions to the prototype,[4] and later to compare users of the MELVYL MEDLINE and CURRENT CONTENTS databases.[5] Users are not forced to take these surveys and are allowed to stop at any time. Response rates are not as high as those for questionnaires administered in person, but they are substantial enough to provide a snapshot of user characteristics and stated reasons for using the databases. Usage statistics reflect overall levels of system use and serve as a kind of Nielsen rating of specific commands and access points. By recording usage patterns for each UC campus, we can also see how the MELVYL system plays in different markets. Statistical reports have grown from a single report for the catalog to an overall system report, detailed reports for each database, statistics on use by individual terminals in each UC library, and figures for incoming use from Internet addresses.[6] These raw data do not tell the whole story, but they have alerted us to problem areas or interesting anomalies for further investigation.

A direct-dial catalog assistance desk at the MELVYL system offices was another original effort to maintain contact with users. Initially, it was used primarily by library staff to report telecommunications and other failures that resulted in the catalog being unavailable. The latest log of what is now called the Helpline revealed dozens of calls not only from UC campuses, but from New York, Texas, Oklahoma, Colorado, and South Dakota. Most were calls for information about how to connect to the system, but they ranged from trouble reports about a transformer fire on the Berkeley campus that prevented that campus from connecting, to analyses of problems with a dozen different communications software packages, to a caller expressing thanks for letting her "mouth off" about a problem. One entry simply read, "Caller had just lots of questions. Lots." Nowadays, the callers are just as likely to be "real users" as library staff. Often they call with a simple request and we are able to direct them to a whole suite of features that meet their needs. Emotions range from anger to euphoria; questions range from technical details to requests for where to find everything everywhere. Even though we do not provide a toll-free number, some callers are happy to pay the charges as long as they can talk to a real person. Even when we cannot solve their problems immediately, most users feel better just being able to report them and have someone listen.

USERS TALK BACK

The tools that accompanied the early development of the system gave us mountains of data about our users, but with the exception of the Helpline, did not give our users many convenient opportunities to talk to us directly. The online version of a suggestion box, the COMMENT command, was implemented in 1983. At first, it was a cautious attempt to give users a voice. DLA staff answered selected comments in print and periodically sent these responses to campuses. It is doubtful whether most users ever found the answers to their comments. Again, DLA benefitted from the information users were providing, but users didn't get a direct response.

Then, in 1986 the Comments database transformed the well-hidden suggestion box into a blank wall beckoning to graffiti artists. All comments, with few exceptions, were posted in a database that users could search using the familiar MELVYL command structure. Soon the MELVYL system had its first fan club, the DILCUEs.[7] One of DLA's rules was that real names would not be posted, an attempt to prevent the system from becoming a message board and to guarantee anonymity as a way of promoting frank comments. Instead of mimicking the "letters to the editor" column where signatures are mandatory, this rule appealed to the more playful and creative commenters who took great pride in devising clever pseudonyms. The DILCUE club and their imitators flooded the database with their exchanges, which often included real questions or anthropomorphic appeals to "MELVYL" to grant them favors. Despite their irreverent comments, they were amusingly reverential toward the system.

Ironically, on the same day that our User Services advisory committee agreed we should stop posting unanswered comments to free staff time for more important tasks, several DILCUE members made a surprise pilgrimage to the DLA offices. They wanted to meet the mysterious woman who answers comments, but since she was not in that day, the mystery was perpetuated. They lined up underneath a sign with "Division of Library Automation" on it and had us take their picture to prove they had been here. It was December and they serenaded us with MELVYLean ballads set to Christmas carols. We in turn rewarded them with a complimentary copy of the MELVYL System Reference Manual, autographed of course! After breaking the news to them about the demise of the free-wheeling comments policy, we entreated them to help us pacify other commentators. They took it in stride and were thrilled when we allowed them to send one last comment direct from the DLA offices. Some of them still send "legitimate" comments.

Some of us have speculated on what this group's existence suggests about the MELVYL system. Do people naturally want to anthropomorphize and personalize a system, and therefore conclude that it is user-friendly, or is the fact that users were able to view the MELVYL system in this personal way a reflection of its user friendliness? Some critics believe the user interface is verbose. Others maintain that the tone strikes the right balance between being overly chatty and cryptically terse by seeming to be generously explanatory. Those who study human-computer interaction do not seem to have the answers either since what is comforting to one user may be annoying to another.

The more "serious" comment database now posts about 200 comments a month. As with the Helpline, topics run the gamut from basic questions on how to conduct a search to detailed suggestions for system improvements. Emotions are expressed here too, as are vastly different perceptions of the system s user friendliness, indicated by these two comments that appeared next to each other in a recent set of postings:

> This is a confusing system!

> You are a wonderful tool to use as opposed to the card catalog.

Another recent comment evaluated the comment feature itself:

> Thanks, MEL, for the opportunity to send comments and get real answers.

Experimental mode is another attempt to get direct feedback from users. Substantial new features or changes are tested here first and users are invited to send comments on their reactions. Sometimes the experiment asks users to comment on specific questions, but usually the experiment description simply explains the purpose of the experiment, how to issue commands, and which help screens to consult.

This process has helped us exterminate bugs and refine features before they go public, but it is like another marketing technique—handing out free samples of a new product at the supermarket. The disadvantage is that the people who may be the most likely users will not necessarily see it during the time period it is available. We do announce experiments, but only regular readers of the system or database news are likely to see announcements. Librarians are encouraged to use experiments, but they often do not want to "risk" trying out changes or new features while they are under the pressure of a busy reference desk. Thus, Experimental mode is not the best way to hear from users, but it is an essential piece of the puzzle.

ADVISE AND CONSENT

User comments, reactions to experiments, and helpline calls ultimately reflect at best a fragmented view of the system and at worst an extremely atypical perspective. An essential complement to these rather raw contributions are the more reasoned and informed discussions we have with our user advisory groups. The Bibliographic Projects Advisory Group (BPAC), a subset of the council of University Librarians from the nine campuses, advises DLA on matters of policy and overall planning. Reporting to them and advising us on specific user issues is the User Services Group. USC consists primarily of reference librarians from the campuses, people who see the system in action every day. Alan Ritch, a former chair of the group, describes USC's tasks this way: "There are organizational and technical changes to announce and evaluate in terms of their effect on the user; experiments to analyze, improve, and implement; instructional ideas and materials to share; MELVYLean mysteries to solve; search problems to puzzle over; search successes to applaud; documentation to design; comments on which to comment; and error messages and HELP screens to be clarified and made more specific."[8]

The group benefits from a mix of long-standing members with a historical view of the system's mission, successes and failures, and new representatives who come with a fresh view, perhaps with experience on other systems, and with new energy and enthusiasm. Sometimes these new members also appear with a few stars in their eyes about how easy or essential changes "should" be. As their understanding of the system deepens and their knowledge of some of the technical underpinnings increases, they come to realize, in evaluating which changes pay back their investment in time and effort. Complacency is not a characteristic to be sought in this group. As Ritch states, We have learned to be neither bashful in our suggestions nor humble in the repetition of old recommendations." Some problems linger, others reappear in cycles or new guises, as the system expands and usage grows. This group's gadfly role keeps us focused on the problems that matter.

The advisory group model was expanded when we began to mount abstracting and indexing databases on the MELVYL system. The first project was the grant-funded MELVYL MEDLINE database. The project called for an advisory group of medical librarians to advise DLA on the special needs of this database and its demanding audience. This smaller group, representing two of the campuses with medical schools, functioned as part of the design team and participated in some of the excruciatingly detailed discussions on how to transform the National Library of Medicine tapes into a MELVYLean image. The group also designed and administered portions of the extensive evaluations that were conducted online and offline, and wrote specifications for statistical reports to be generated by the system.

This model was so successful that it was continued almost without question as we moved on to the next databases. Now a task force is appointed to assist with the initial implementation of article databases. Members are selected for their expertise in the subject area or their teaching and training experience. After the database becomes fully operational, the task force is dissolved and continuing issues are handled by the User Services Group.

One aspect of the MELVYL MEDLINE model that was not continued was the appointment of a faculty and graduate student advisory group. For MEDLINE, this group of experienced MEDLINE users representing different levels of teaching and research provided a very different view of what worked, what was essential, and what was not important. Their contributions were invaluable, but the logistics of convening a similar group for subsequent databases argued against the automatic appointment of such a group for every database project. The next two database projects, CURRENT CONTENTS and the Information Access Company databases (Expanded Academic Index, National Newspaper Index, and Computer Database), were interdisciplinary in appeal and did not have as much of an established base of users amongst the faculty, making it more difficult to identify a group of experts.

As the process for selecting databases and other electronic resources has evolved and become more established, we have benefitted from the advice of another group, the Computer Files Committee. This smaller group does not have a representative from each campus but rather from each functional area of

library operations that is concerned with the selection and service issues. Members represent collection development, public services administration, and front-line reference and database services librarians. This group surveys campus librarians for information concerning current CD-ROM and print subscriptions, specific database recommendations, and general priorities for electronic access. They have outlined general strategies and policies as well as evaluated detailed offers from database vendors.

One of the real drawbacks of this system of advisory groups is the time it adds to the process. Convening representatives from across the state for face-to-face meetings is both time consuming and expensive. The groups usually meet once with DLA staff to review the initial project plan and discuss major implementation issues. In later phases of the project, DLA often needs to make decisions quickly as new information is uncovered during the course of analysis and programming. Fortunately, most librarians now have access to electronic mail, and we have been able to conduct quick polls and much of our other business via special mail discussion lists. Having the discussions online also helps representatives share the background and rationale for decisions with their colleagues.

The pauses in the schedule to allow advisory groups time to ponder are well worth it in the long run. Not only do we often gain from their subject expertise or good ideas, we also gain from their support of a project they have helped to shape and ultimately will have to teach and promote to the user community.

The advisory groups have another outlet for expressing their opinions about the system. An informal, monthly newsletter, *The Mynd of the MELVYL System* (known as *MOM*), functions as a sort of "Consumer Reports" as it evaluates new "products," makes recommendations on using them, discovers flaws, and generally rates the system's performance.[9]

We know we have to be on our toes with *MOM* always watching over our shoulder.

GREAT EXPECTATIONS

Even with all of these different means for learning about our users and learning from them, we still have a long way to go toward meeting their needs. changed over the years. Our expectations about users have changed over the years. The original designers stated that "users can teach themselves the command language simply by using the HELP commands," and we still hope that the system is largely self-teaching. But we no longer expect users to absorb the full power and complexity of the system on their own, much less understand the intricate system of cataloging upon which it is built. While it was anticipated that campuses would do some demonstrations of the system, it was probably not envisioned that teaching and training would occupy so much of the librarians' time through the years.

Similarly, the expectations of users have changed. In the early days, not only were online catalogs uncommon, but personal computers were just beginning to proliferate. New users were more easily impressed with the system simply because it was a novelty. For more experienced users, the MELVYL system might have been at the center of their bibliographic universe until the last few years, providing the only easily accessible electronic access to bibliographic information. Now it is only one of many screens beckoning from reference areas crowded with terminals, most of which are much sleeker than the aging terminals dedicated to the MELVYL system. Alternately, it may be presented as one of many options from a single terminal, a complementary system to the campus's local catalog or a gateway to other systems on the Internet that may be more compelling for a given user's subject area. As remote access through the Internet has increased, a new class of users outside the UC community is viewing the MELVYL system as just another stop along the way, to be consulted occasionally only if it has something special to offer. Many users are more computer literate than the first eager users of the system. They constantly compare it to other systems they have used, and often wonder why it can't perform some tasks as nimbly as their own personal computers.

The MELVYL system can claim only so much of users' time and loyalty, and must become even "friendlier" to attract the attention of the ever-changing parade of new users. It is no wonder that the casual commenter, the helpline caller, and the advisory group member all cajole us to make the MELVYL system go faster, grow even bigger, and extend its boundaries even farther.

REFERENCES AND NOTES

1. Barbara S. Radke, Katharina E. Klemperer, and Michael G. Berger, "The User-Friendly Catalog: Patron Access to MELVYL," *Information Technology and Libraries* 1, no.4:358 (Dec. 1982).
2. Ibid., p.360.
3. Ibid., p.358.
4. *University of California Users Look at MELVYL: Results of a Survey of Users of the University of California Prototype Online Union Catalog: Final report to the Council on Library Resources* [Appendixes] (Berkeley, Calif.: Division of Library Automation and Library Studies and Research Division, Office of the Assistant Vice-President—Library Plans and Policies, University of California Systemwide Administration, 1983).
5. For descriptions of the questionnaires and results see Margaret Gordon, "Users of the MELVYL Current Contents Database: A Survey and Evaluation," *DLA Bulletin* 11, no.1:14–16 (Summer 1991); Mary Horres, Susan Starr, and Laine Farley, "Evaluation of the MELVYL MEDLINE Database and Interface," *DLA Bulletin* 9, no.2:14–16 (Fall 1989); and Mary Horres, Susan Starr, and Beverly Renford, "MELVYL MEDLINE: A Library Services Perspective, *Bulletin of the Medical Library Association* 79, no.3:309–320 (July 1991).
6. To see MELVYL system statistics, TELNET to MELVYL.UCOP.EDU or 31.0.0.13, and enter SHOW STATISTICS for a directory of statistics available.
7. The DILCUE commenters revealed that "DlLCUE" is simply "Euclid" spelled backwards.
8. Alan Ritch, "The MELVYL Catalog User Services Group: Shaping the Catalog," *DLA Bulletin* 7, no.2:20 (1987).
9. To see the latest issues of the *Mynd of the MELVYL System*, TELNET to MELVYL.UCOP.EDU or 31.0.0.13 and enter SHOW MOM. ■■

Computing Resources for an Online Catalog: Ten Years Later

Mark Needleman

In this article I will look at the changing computing resources required to run a large information retrieval system like the MELVYL system. I will give a historical perspective, examining the early days of the MELVYL system and how the system has changed in the last decade. I will also provide some perspective on how new technologies have affected large information retrieval systems and what technologies will be required of such systems in the future, as the demands and requirements of large information retrieval systems continue to mature to meet increasing user needs. This article updates a 1982 ITAL *article on the same subject.*

The original prototype of the MELVYL system came online in 1981, as a nonupdatable database of approximately 750,000 records accessible through approximately nine terminals on each of the nine campuses of the University of California. These terminals were hard wired into the MELVYL system. In 1983, the production MELVYL system came online, consisting of an updatable database of the UC catalogs and periodicals holdings. The original hardware environment included four IBM-compatible mainframes, each capable of executing approximately one million instructions per second. This was later upgraded to three IBM 4381 mainframe computers, each capable of about 2.5 million instructions per second. Two of the machines were used for the MELVYL catalog, while the third 4381 was used for development and loading the database. The older IBM-compatible machine was kept as a test environment. These machines ran the OS/360 MVT IBM operating system, which was a real memory operating system limited to 16 megabytes of memory. The operating system, the MELVYL program, and all of its supporting programs, such as telecommunications software, had to fit in those 16 megabytes of memory. At its peak, this configuration was capable of supporting fifty simultaneous users.

To put that environment in perspective, today that amount of memory is commonly

Mark Needleman is Coordinator of Advanced Technology Development, University of California, Office of the President, Oakland, California.

found in average-sized UNIX-based desktop workstations. It is not uncommon to find PCs containing eight megabytes (or more) of memory—and these are essentially single-use machines. In sharp contrast, today the MELVYL system exists on a large IBM 3090 mainframe computer that executes approximately forty million instructions per second. About 250 megabytes of memory are available on the machine, and the operating system uses an architecture known as virtual memory that limits neither the size of a program running on it nor the size of the physical memory of the machine. In fact, the MELVYL system and each of its supporting programs have available to them up to two gigabytes of this virtual memory space. Recently peak usage periods have seen more than five hundred simultaneous users on the system. That number refers to users actively employing the system to perform searches and display records. There can also be several hundred more idle terminals that are connected to the system but are not actively using it. The system, especially the networking software, must have the ability to support those added idle connections.

Another important area that reflects how things have changed is that of disk technology. When the system first went online, the disk drives used for storing the database stored either 200 megabytes or 300 megabytes each. The total disk storage available was around 20 gigabytes, and the disks had an average access time of around 30–40 milliseconds. Today's disks store close to two gigabytes each yet have an average access time of around 15 milliseconds—an 8 to 10 percent increase in storage capacity with a corresponding decrease in access time by better than 50 percent. The disk drives also use technology that allows frequently accessed data to be made available without the necessity of actually reading it off the physical media, even further reducing the access time in many cases. Currently, the MELVYL system has more than 150 gigabytes of disk storage in use. This increase is required by both an increase in the size of the bibliographic database and a change in its role, having grown from just an online catalog to a system that also provides access to multiple abstracting and indexing (A&I) databases.

One final comment on these changes in the hardware base for the MELVYL system: In the early days, hardware failures were not uncommon and were a major cause of down time for the system. Disk drives would lose information that would have to be restored from backup tapes, or the computers themselves would experience failures, necessitating visits from repair personnel. This could cause protracted outages. With today's hardware base, it is not uncommon for the mainframe computer to run for several years without a hardware failure. Further, even if there is a failure of some kind, modern mainframes have redundant systems built into them that allow them to keep operating without needing to be taken down. Repairs can thus be scheduled for off hours. This is also true of some of the the new disk technologies. Failures are few and far between. The newer operating systems now in use are also extremely reliable. The software has the ability to detect and survive failures in many of its components, allowing it to essentially bypass or route around the failing components. All of this has led to an up-time rate for the MELVYL system that usually exceeds 99.9 percent; this includes the time that the system is brought down for scheduled hardware and software maintenance.

However, the area that has shown the most dramatic changes in the life of the MELVYL system has been the telecommunications environment. The original telecommunications network, when initially deployed, consisted of about nine hard-wired terminals at each of the nine University of California campuses. These terminals, located in the libraries, ran, at best, at 1200 bpi through a single multiplexed phone line back to the computer center in which the MELVYL system ran. There was no redundancy or protection against failure in the system. The loss of the multiplexor or phone line would take out all access to a campus, and of course, the only access to the system was by going to the library to use one of the few available terminals.

It was obvious from the beginning that this setup would not support the large-scale information retrieval infrastructure we were attempting to build with the MELVYL system, and we soon replaced it with a telecommunications network based on the TCP/IP Internet protocols. Initially, the network consisted of a single node on each campus, with terminals in the libraries using that node

to communicate back to the central site. Over time, the network has been upgraded both in the speeds of the various lines and in the connections among the nodes, so that there are now multiple links among the nodes with redundant paths to protect against individual line loss causing loss of access for any campus. We have also connected the MELVYL point of presence on each campus to the network backbone that exists on each campus, so that any terminal or host that has access to the campus network can gain access to the MELVYL system using the TCP/IP and TELNET protocols. Because the campus networks are connected to the regional networks in the state of California, and in turn those regional networks are connected to national and international backbones, the MELVYL system is accessible from virtually anywhere on the international TCP/IP Internet. Gone are the days when users were required to travel to the university libraries to gain access. Now students, faculty, and staff can access the system from any location on campus and beyond that has access to the campus networks. The terminal population that can access the MELVYL system has grown from the initial eight hardwired terminals to one that literally encompasses tens of thousands of terminals and workstations. I believe that this has been the most profound development in the technology of the MELVYL system since its inception and the one with the broadest implications for its future development and for the future development and changing role of information retrieval systems in general.

GENERAL OBSERVATIONS

Having discussed some of the changes that have occurred in the technology base of the MELVYL system in the years since its initial deployment, it might be interesting to make some observations about what all the statistics mean, how the current MELVYL system differs from its original incarnation, and what all this portends for the future.

First, usage of the MELVYL system has increased dramatically in the last decade. Today, during peak usage periods, there are more than five hundred simultaneous users on the system. These users perform more than 500,000 queries a week, displaying close to four million records. Also, the scope and coverage of the system have grown in the last decade. From an initial single database of 500,000 bibliographic records it has grown to a system consisting of multiple databases that together contain more than thirteen million records. Besides the basic bibliographic files, there are several A&I databases providing coverage of a wide variety of subject areas, with many more to be added over the next several years. The system also serves as a gateway outbound to many other information systems on the network. This increased usage and scope would have been impossible without the increased technology base that has been added to the system over the years.

Second, the sophistication and levels of functionality of these systems have grown immensely in the last decade. Systems like the MELVYL system continue to offer a growing array of services and features to their users. Much of this is in response to greater user sophistication and demand. Again, much of this would not be possible without the corresponding increase in technology that these systems have seen in the last decade. A further growth in such technology is going to be required in the coming years to meet ever-increasing user requirements.

Third, the growth of networks and network-available information resources have exploded in the last decade and will continue to make incredible strides in the future. The network component of information retrieval systems is a growing portion of the totality of the services they provide. The era of standalone information retrieval systems is ending, and systems that cannot adapt and use networks will become increasingly irrelevant in the future. The entire aspect of networked-based information services is changing the paradigm underlying how information systems are designed, deployed, and used. It is important that this model be built in as our current information retrieval systems are expanded and adapted to meet future needs and provide new services. Many of today's generation of library systems are integrated systems in which all automated library functions are incorporated. It will be interesting to watch the effects the growth of networks will have on such systems, now that it will become less necessary to provide all such functions in a single system and increasingly cost-effective to use the network. The network will offer the opportunity to choose technology that is best for a particular function, removing the neces-

sity of having to compromise on one technology for all functions. A possible outcome of the tremendous network growth may be the demise of the integrated library system model.

Finally, technology has changed dramatically since the early days of the MELVYL system. When the system was initially built, the personal computer had just been introduced, and no one yet had any idea about the kind of computing revolution it would spawn. Macintoshes did not yet exist, and the graphical and mouse-based interfaces they would popularize could only be found in advanced research labs. The desktop UNIX workstation was still in the design stage.

A basic challenge for today's generation of information retrieval systems will be to incorporate this technology in a meaningful and productive manner. Tremendous computing resources exist on these systems, and information retrieval systems will need to find good ways to exploit such resources. They open up the possibilities of new searching techniques and new user interfaces, and will move beyond the current world of structured hierarchical data to one that includes full text and nontextual data, such as images and audio and visual material. To accomplish this, today's systems will need to move beyond their current monolithic designs to employ new client/server technology that can incorporate these workstations as full partners in the information retrieval process.

Within the MELVYL system, such a project is currently in the design stage. The plan over the next several years is to replace the current design with one based on a client/server architecture. The user interface portion of the system will migrate to a UNIX-based platform. Both a line-oriented interface, similar to the current one, and a graphical user interface using the X-Windows protocol will be supported. The IBM 3090 mainframe will be converted to a database server that will be accessed by those interfaces using the U.S. ANSI standard Z39.50 information retrieval protocol. Because we feel that there currently is limited understanding of what a graphical user interface for information retrieval systems should look like, we expect to go through several iterations of prototyping such an interface to gain a better understanding of what features make sense and best meet user needs. Because Z39.50 provides a standardized mechanism for accessing the MELVYL system databases, we expect to see the development of many other user interfaces for a wide variety of platforms, allowing users to choose and customize the interface that best suits their needs. Our expectation is that a similar design strategy will be developed by many of the current generation of information retrieval systems.

CONCLUSIONS

This article has provided an update on the computing environment on which the MELVYL system runs and how that environment has changed and grown since its initial deployment. It has discussed some of the factors and requirements that have driven that growth and speculated about what kinds of technologies will be required to support the future growth of such systems. While it is true that new and advanced technologies have allowed such systems to grow, it is perhaps even more true that ever-increasing user requirements, demands, and expectations will drive the need for new technology in the future.

BIBLIOGRAPHY

Arms, Caroline, ed. *Campus Strategies for Libraries and Electronic Information.* EDUCOM Strategies Series on Information Technology. Bedford, Mass.: Digital Pr., 1990.

Arms, William Y., and Thomas J. Michalak. "Carnegie Mellon University." *In Campus Strategies for Libraries and Electronic Information.* Ed. Caroline Arms. EDUCOM Strategies Series on Information Technology. Bedford, Mass.: Digital Pr., 1990. P.243–71.

Bearman, Toni Carbo. "The Changing Needs of Users of Scientific, Technical, and Societal Information." Paper presented at the International Symposium on the Future of Scientific, Technological, and Industrial Information Service, Leningrad, USSR, May 1990.

Lynch, Clifford A. "From Telecommunications to Networking: The MELVYL Online Union Catalog and the Development of Intercampus Networks at the University of California." *Library Hi Tech* 7, no.2:61–83 (Issue 26, 1989).

———. "The Z39.50 Information Retrieval Protocol." *Computer Communication Review (Sigcomm)* 21, no.1:58–70 (Jan. 1991).

———. "Networked Information: A Revolution in Progress." *Proceedings of the Lancaster Clinic: Networks, Open Access, and Virtual*

Libraries: Implications for the Research Library. Urbana, Ill: Graduate School of Library and Information Science, University of Illinois, in press.

Lynch, Clifford A., Mary E. Engle, and Mark H. Needleman. "Computing Resources for an Online Catalog." *Information Technology and Libraries* 2, no.1:58–72 (Mar. 1983).

Needleman, Mark H. "The Z39.50 Protocol: An Implementor's Perspective." In *Resource Sharing and Information Networks.* In press. ■■

Ten Years of Monitoring MELVYL: A Librarian's View

Alan Ritch

The first ten years of the MELVYL system have profoundly affected the lives of University of California librarians. The rapid growth of the system's content, complexity, and use has required frequent modifications of its interface. These changes have required the continuous involvement of librarians in advising the system's designers on new features and new databases, in instructing users, and in observing user behavior. This article traces, from a librarian's perspective, the evolution of the system from its origins as a powerful prototype online catalog to its present role as a complex of multiple databases, services, and resources. The article's primary focus is on ways in which the growth, refinement, and development of the system have entailed adaptive design, flexible instruction, and user tolerance for change.

ON FIRST SEEING UCPOLUC IN THE SPRING OF 1981

Ten years ago there were still fruit orchards in Northern California's Silicon Valley, now populated only by plants, not to mention apples, of a different kind. They were fragrantly and picturesquely in bloom on the March day carloads of librarians passed through on their way to West Valley College, Saratoga, for a workshop on bibliographic instruction, innocuously entitled "Teaching the Changing Catalog." I was there looking for new ways to teach the students and faculty of the University of California at Santa Cruz how to cope with the transition from a book catalog to a microfiche catalog. I was vaguely interested in the promised presentation by one Katharina Klemperer from the UC Division of Library Automation (DLA): "Teaching the Use of a Prototype, Online User-Friendly Catalog," evidently a preview of a forthcoming attraction. Obviously this cumbersome label—the University of California Prototype Online Union Catalog—badly needed an acronym. Little did we know that the name given to the strange new system, one more mystifying but more mellifluous than the expected UCPOLUC, would become a mantra for users of UC libraries. *MELVYL* has evolved from noun to adjective, from catalog to system, from tool to paradigm, and many of our professional lives have evolved with it, few more symbiotically or parasitically than my own.

West Valley College in March 1981 already, precociously, had its own online catalog, a glittering new medium with a modestly familiar message. Its simple menu of choices reassuringly echoed the search options of the card, allowing the novice to search for books by author, title, or subject. But when Klemperer, with Mike Berger, another member of the UC DLA team, gave her presentation on teaching the new UC catalog, the audience was instantly energized, not just by persuasive pedagogy but by another kind of power. Their teaching tool was a primitive Texas Instruments terminal with no screen. (This was, after all, very early in the year that saw the birth not just of the MELVYL system but of the IBM PC.) However, we became suddenly aware of a message that fully exploited the online medium. We were astonished to discover that online technology fostered a new breed of catalog, one that allowed virtually every word in the record to be retrieved, separately or in Boolean combination, that provided two modes of searching according to the user's experience, that delivered

Alan Ritch is Social Sciences Bibliographer, McHenry Library, University of California, Santa Cruz.

user assistance directly to the terminal through well-written help screens, that unified the holdings of all nine UC campuses, and that potentially standardized the most important tool for bibliographic access in UC libraries. As an instruction librarian, I was utterly exhilarated by the prospect. Suddenly there was much more to learn, much more to teach, and much more with which to teach.

COOPERATIVE ENTERPRISE: "NOT AN EVENT BUT A PROCESS"

I was even more enthralled by the prospect of a productively cooperative enterprise between the evidently approachable designers of the MELVYL catalog and its users—professional and nonprofessional, scholarly and casual—and among the librarians of the far flung outposts of the university. As a junior librarian in the smallest general library of the UC empire, I identified an irresistible opportunity for involvement in the most important outcome of the visionary Salmon Plan, conceived four years earlier as a blueprint for the unification of UC libraries, effecting their transformation from some of the best public university research libraries in the country to one of the greatest research libraries in the world. From being a small fish in a small pond, I could become a small fish in the much larger Salmon hatchery.

My naively optimistic expectations were founded on the assumption that there was now a common catalog, far more standardized than the card catalog and far more capable of teaching itself than the old cardware had been. I assumed that there would need to be explanatory handouts, exercises, fliers, and flipguides to supplement the online instruction; that the design and production of these could be centralized; that the DLA staff themselves would be thoroughly occupied with more technically demanding activities; and that a few enterprising librarians could develop a universitywide clearinghouse of materials literally once and for all.

Obviously I overestimated the stability of a system, whose arrival, we were warned from the outset, was "not an event but a process"; I overestimated the new unity of UC libraries; and I overestimated the catalog's identity as a common system, wherever it was placed. Its dynamics have precluded permanent instructional solutions. Its increasing scale and diversity and an unexpectedly large volume of use have required continuous refinement of the user interface, changes that have required alertness among all users, but especially those responsible for promoting innovation. UC unity has been undermined by the creation, at several of the nine campuses, of a series of local online systems, with additional features, such as circulation status and call number searching. Such local integrated systems have had, from the universitywide perspective, a disintegrating effect. Seen by some as complementary and by others as competitive, they have been admitted even by their installers to be depressingly expensive during a period of diminishing resources. This same independence of purpose, this multiple institutional identity has given the MELVYL system itself multiple roles that vary according to the campus context.

A recent expression of this diversity is the relative use of the CURRENT CONTENTS database (CC), an index to 6,500 scholarly journals in all fields loaded on the MELVYL system two years ago. This resource, presumably of universal interest, has been virtually ignored in locations where the MELVYL catalog has been underexploited. UCLA's use of CC in May 1991 was one-third that of the smaller UC Davis campus. Even little UC Santa Cruz had higher use figures than giant Los Angeles. Attempts to preach the MELVYL system message have depended, with mixed success, on local missionaries and local translations of the scripture.

COPING WITH INSTANT AND CONSTANT OBSOLESCENCE

The earliest instructional materials inevitably flowed from the system source. DLA set a daunting example of graphic clarity and instructional thoroughness in its first user guide, printed in August 1981. This first edition was beautifully designed, crisply folded into fifteen panels, difficult to cannibalize, and almost instantly obsolete. I still have a copy of this handsome artifact of MELVYL's origins. Its examples, instructive in 1981, are still instructive, but for disconcertingly different reasons. Consider the following sample searches from the pamphlet entitled, with unintentional irony, "Quick Guide":

FIND SU POWER AND NOT EXECUTIVE

FIND TI MAGIC OR SU MAGIC

BROWSE PA MILLER

BROWSE SU ENERGY AND NOT SOLAR

Each of these examples was perfectly reasonable in the static prototype database of about 730,000 records being searched by a few inexperienced but skillful pioneers. They are no longer exemplary searches. Indeed, not one of them is even allowed during periods of peak load: the search terms are too common to be easily searchable in a database that has become astonishingly big and busy during the intervening decade. The catalog database is roughly ten times as large as the prototype was and the number of its simultaneous users often exceeds four hundred on active afternoons. The old model searches illustrated above were not simply sensible but reveal considerable sophistication in their use of Boolean logic. Now, however, their terms are so common that they would take an unacceptably long time to process. They would also unacceptably retard hundreds of other simultaneous searches. As a result, from noon to 5 p.m. they are simply disallowed, in order to fulfill the utilitarian goal of the greatest good for the greatest number.

Other "quick" examples from the 1981 guide were similarly overtaken by growth both of the database and its user population:

AND SU ALCOHOL

BROWSE TI LIBR# AUTOMATI#

The first, a sensible Boolean add-on modification in many search contexts, is now a "long search." Long searches trigger special warning screens that discourage the user from continuing and encourage the reformulation of the search. For example, a user who enters

FIND SU SEX BEHAVIOR

and thereby discovers more than seven hundred items may be smart enough to recognize that this is too many to scan, and may even be smart enough to add on

AND SU ALCOHOL

to limit the result to items concerned with both topics. This smart user will be discouraged from continuing the search and urged to see a reference librarian for suggestions. The poor reference librarian, also enduring the busiest part of the day, is more likely to be impressed by the conceptual sophistication of the search than capable of improving it. In fact, if this long search is patiently continued to its outcome, it actually saves the user a much longer visual search through hundreds of titles and does yield a handful of the most relevant items. Few UC librarians can explain and even fewer try to justify these processing difficulties, which stem from the MELVYL system's inability to search for a second term within a set that has already been retrieved. Instead of searching in a set of several hundred, it must return to the entire database of several million before making the Boolean linkage.

The second example, which illustrates the browsing of (heavily truncated) terms in titles, was sacrificed in the production catalog to the sensible goal of sustainable system speed. In addition, the BROWSE feature, which allows the user to scan lists of numbered headings and to SELECT headings for searching, was not installed for any of the indexes until 1985. It has never been restored to title searching, and it has never been widely or enthusiastically exploited, even in the subject index. We librarians have not actively promoted it, partly because such reasonable searches as

BROWSE SU SANTA CRUZ

are obstructed, first by a long search message and then, if that warning is (guiltily) ignored, by the system's refusal to display the subdivisions for the most common headings, the very headings for which the display of subheadings would be most useful!

PERFORMANCE PROBLEMS

I intend here no facile criticism of the shortsightedness of the designers of an otherwise elegant and useful handout or indeed of those who designed an otherwise magnificent and invaluable resource. Although the eventual size of the database could have been estimated fairly precisely, the volume and nature of user demand for what was originally conceived as a union catalog could not have been foreseen, nor could the destructive interaction of these relentless spirals of growth.

Expectations about response time were based on the unrealistic experience of the prototype or on falsely analogous experiences with smaller, less elaborate systems, or with larger databases more generously supported by disk space and hardware. The prototype was tantalizingly liberal in its encouragement

of the user to retrieve the most comprehensive possible results. The librarians who were scrutinizing the catalog's early features tended to be very demanding searchers who had cut their teeth on DIALOG and BRS and wanted the same strength and flexibility in the new resource. Naively, we all assumed that the powerful features that worked so well for experienced explorers of very large databases would work equally well for those completely innocent of the scale of the search environment, the significance of subject headings, and the syntax and grammar of retrieval. Those of us who were more pessimistic about instant omniscience hoped that the varieties of ignorance would somehow offset one another. Overly broad searches would be reduced by the insertion of the wrong or too many words; overly precise searches would be redeemed by the casual use of a Boolean OR. Murphy's Laws dictated quite opposite outcomes. System statistics indicated a very high frequency of very large results, very long searches, or searches yielding absolutely nothing; transaction logs showed sequences of stubborn repetitions, misunderstandings, and unproductive decisions; and unobtrusive observation of terminal screens near the reference desks confirmed that an unbearably high number of searches were fraught with evident, avoidable difficulties.

Other early surveys yielded much more optimistic conclusions. The rosy results of the landmark Council on Library Resources (CLR) study on the attitudes of users of online catalogs suggested that virtually everyone loved the new tool, even (perhaps especially?) those who had very little experience with it. One of the surprising discoveries that stemmed from the survey was the unexpectedly high number of subject searches. Perhaps the subject approach had always been more significant than our persistent attachment to the description of known items and their arrangement by main entry would acknowledge. The flexibility of the new systems has made subject searching much more productive than it used to be, at once potentially more rewarding in its results and more challenging in its complexity to both user and system performance.

If there was universal appreciation of the online catalog's performance in its infancy, there was far more ambivalence in the following years, when response time and performance began to be treated as interchangeable attributes. Performance came to be implicitly defined not by what the catalog did but how quickly it did it. Innumerable design decisions in the mid-1980s were driven by an understandable apprehension about the most visible of the system's shortcomings. The emergence at our two largest campuses, UCLA and Berkeley, of two integrated online catalogs, ORION and GLADIS, more modest as retrieval tools but demonstrably more nimble in the tasks they could do, increased the pressure to curtail features, however useful, that were particularly burdensome on the older, larger MELVYL system.

A DECADE OF DRAMA

Although in most cases such modification streamlined searching, not just for the computer but for the user, the process of simplification has been complex, a drawn-out drama equally rewarding for both embroiled participants and attentive witnesses. On the one hand, there were the forces of growth and development in numbers of records and users, formats, and databases; on the other, a heroic band of decison makers, designers, and programmers, nudged and needled by library administrators, user groups, and ordinary users sending an extraordinary volume of messages via the COMMENT feature. That the system should survive such a potentially contentious confrontation is remarkable. That it should thrive is cause for celebration and gratitude.

The MELVYL system's first ten years have been much too prolix, full of loose ends, contradictions, cycles, and reversals, to allow the insertion of tidy chapter headings or restful intermissions. Here is a summary outline of the MELVYLean chronicles, from one librarian's perspective.

In the beginning was the prototype, which raised our expectations about the arrival of a new age, a new language, a new freedom of information retrieval. The arrival of the production catalog ushered in the first period of retrenchment, in which several features were curtailed. An important parallel database of periodical titles was soon added, the California Academic Libraries List of Serials (CALLS). Along with CALLS, however, came a growing chorus of complaints about the rapidly deteriorating response time in the cat-

alog itself, which led to a long series of adaptive measures that either curtailed functionality or added new precision. For example, the original SUBJECT index covered any keyword in either the title or the subject field. This liberality was checked, after less than a year of the production catalog, so that SUBJECT found only words in subject headings. In the following months, the EXACT TITLE and EXACT SUBJECT indexes were introduced, allowing those who had accurate references or who knew actual subject headings to do quick, precise searches. The responsiveness of the system's managers to suggestions gathered from library administrators, user groups, and the increasingly indignant COMMENT medium was in desperate contrast to the system's own ability to respond to increasing user demand. The changes were so frequent and their implications so crucial to effective searching that a regular newsletter, *The Mynd of MELVYL* (*MOM*) was invented, written, and published by librarians to keep users and their system literally up to speed. *MOM,* now approaching its 100th issue, continues to serve this function of grassroots commentary by and for ordinary users.

A telling example of the uncomfortable tension between the theoretical value of an innovation and its operational destructiveness was the ADDED SINCE feature, introduced as a way of coping with increasingly large results by allowing users to limit their retrievals to items added since they last used the database. The actual use of this feature, however, was so cumbersome and time-consuming that after little more than a year, it was radically curtailed. Much more successful was the SMART TITLE feature, which allowed the system to infer the user's intentions from the kind of words entered in the title index and to process it either as an exact title or a keyword search. Intended to save the system unnecessary work, it also, more often than not, saved the user an unnecessarily large result.

Other user aids developed during this critical and challenging phase in the mid-1980s were such features as PARTIAL, which allowed the searcher receiving a "long search" message to retrieve a sample of two hundred records; SAVE, which allowed searchers to save the most relevant items to a list that could then be printed or downloaded; the automatic appearance of the prompt to try a TITLE WORD search for terms that found nothing in the subject heading field; and the FORM limit, which allowed the refinement of a large result to a specified format of material, for example scores, recordings, video, or software. The last item became particularly useful as more and more nonprint media records were loaded into the catalog database.

INSTRUCTIONAL OPPORTUNITIES AND CHALLENGES

These changes and challenges to the interface and the primary database required the almost continuous revision of handouts and instructional strategies. A complex of parallel promotional and teaching activities began in 1987, with the arrival of MEDLINE, the first of a series of additional MELVYL databases, similar in interface but sufficiently different in their structure to require the mastery of new vocabulary (*MeSH*), new concepts (hierarchical TREES), and new strategies (for example, the use of EXPLODED and MAJOR headings). If MEDLINE's features were significantly richer than those of the original catalog, the character of the next database to be added was relatively spartan. CURRENT CONTENTS, lacking any subject fields, could be searched effectively only by those aware of and capable of exploiting keywords in the titles of articles. CC's lack of a thesaurus and controlled vocabulary demanded that its users enter multiple alternate terms linked with those same Boolean conjunctions whose use had not been very successful in the MELVYL catalog itself since its infancy. In 1991 a trio of databases created by Information Access (IAC), Expanded Academic Index, National Newspaper Index, and an index to articles on computers were added to the MELVYL system. Their success was immediate, partly because the journals and newspapers they indexed were of more obvious value to most users, but also because the record structure and search vocabulary were so similar to those of the catalog itself that relatively little user training was needed.

Though these contrasts among the several MELVYL databases proved problematic to ordinary users, they offered to those of us who were active in reference service and library instruction a wonderful opportunity to learn and teach an enormous range of online search opportunities, held together by a common

command language. What was most strongly reinforced by these activities was the need for conceptual generalization and illustrative simplification. Detailed manuals on each database were abridged to flipcharts, pamphlets, bookmarks, and a matrix contrasting all the databases. Such reductiveness was even more essential in teaching the use of the unfamiliar or idiosyncratic interfaces of other OPACs made available to MELVYL system users. Even before CC was loaded, the command USE CARL linked the dumb terminals in any UC library to most of the databases available to the Colorado Alliance of Research Libraries. During the last two years, dozens of other systems have been linked by USE and the Internet to the MELVYL system, presenting new opportunities for promotion, instruction, and confusion.

Not that the users themselves have demanded our assistance. From the CLR surveys of the early 1980s, we discovered that most users learned (or claimed to learn) the system by themselves in fifteen minutes or less. This attitude persists, reflecting, the sceptic would argue, less competence than complacency. The ultimate users of the MELVYL system are a diverse and demanding constituency, impatient for results and indifferent to process. Most do not have or will not give us the time to teach them all the details, and the content of our instructional sessions has become responsively parsimonious. At first, to those few who would listen, we would describe in detail the menus and command language that seemed to echo the mental models and search habits of those we presumed to be familiar with the card catalog. Then, as the context became more computerized, we based our instruction on those features that had no precedent in the earlier technology, e.g., keyword searching, conjunctions, the ability and need to balance recall and precision. Most recently, as more and more data and databases have become accessible, teaching the techniques of access has required even more radical simplification to allow time to teach multiple search dialects, to foster discrimination among sources, and especially to reverse the common assumption that minimally described, remotely located items are always superior to those held locally.

Perhaps our most important instructional function is the promotion of new features. The great success of MELVYL MEDLINE and the immediate enthusiasm that greeted the trio of Information Access databases stemmed from their self-evident value for those already very familiar with their virtues via other media. MELVYL CURRENT CONTENTS, in spite of its virtually universal subject scope, has been adopted much more slowly. Evidently its discovery required far more skillful and intensive marketing campaigns than we have so far been capable of mounting. The creation in 1990 of a small subset of recently published materials has been, surprisingly, similarly ignored by most users, in spite of its potential advantages in speed, flexibility, and even, for many fields, relevance over the parent catalog. The effort and expense of developing this database, known as TEN because it includes materials in UC libraries published in the last ten years, has yet to be rewarded by widespread adoption. There is some irony that the system's ten-year anniversary should be marred by the failure of 95 percent of its users to discover the "database of the decade." Perhaps if we tried this phrase as a marketing slogan, we might see more use of this neglected catalog of the 1 1/2 million records of items published in the decade since the birth of the MELVYL system.

THE ENDLESS EXPERIMENT

If we librarians have been sometimes frustrated and often frazzled by the challenge of change, we must also, to a remarkable extent, take responsibility for the system's mutability, since DLA has always listened respectfully and perhaps too attentively to the clamor of its noisiest users. This hospitality to advice and the reliably restless creativity of the DLA staff have, from the beginning, made of the MELVYL system an endless experiment. Eventually, halfway through the decade, a formal Experimental Mode was invented as a sensible launching pad for new features. That event (or process) roughly coincided with the adoption and swift modification of such flawed innovations as the ADDED SINCE feature, which, although never a formal experiment, lasted in its liberal form for only fourteen months. During the last five years, more than forty new features have been introduced in the Experimental Mode. Most, sometimes modified by user suggestions, have eventually been adopted.

We look forward to future experiments to test future innovations, such as new databases, their interfaces, new facilities to increase the system's effectiveness as a bridge between the growing population of remote users and the growing number and diversity of remote resources, and new modes of downloading and document delivery. In spite of the turmoil that surrounded the MELVYL system and the origins of the formal Experimental Mode in the mid-1980s, I hesitate to characterize that period as a midlife crisis, since that metaphor would imply current decrepitude and imminent demise. Nevertheless, in celebrating important anniversaries, life-cycle references are inevitable. Perhaps the MELVYL system is truly a boisterous ten-year old, still growing, still tasting and testing everything within reach, still groping for new toys just out of reach, surprisingly full of wisdom and experience, but still having much to learn and new information worlds to conquer. ■■

APPENDIX A
A CHRONOLOGY OF THE MELVYL SYSTEM: TEN YEARS OF CHANGES

3/81	Demonstration of the University of California online union catalog.
4/81	Prototype catalog available to UC library staff.
6/81	"MELVYL" demonstrated at ALA Conference, San Francisco.
8/81	MELVYL Prototype widely available.
3/82	Online survey of MELVYL users.
11/82	*Users Look at Online Catalogs.* Council on Library Resources Report.
8/83	User comment accepted online.
11/83	Production MELVYL Catalog released.
2/84	Mynd of MELVYL (MOM) first issued at UCB.
4/84	Prototype catalog retired.
5/84	1 million titles in MELVYL Catalog
8/84	California Academic Libraries List of Serials (CALLS) on MELVYL System.
10/84	DATE limits. Interface changes to improve response-time: SUBJECT index no longer covers title keywords; no more single letter truncation.
11/84	Exact Title searching introduced.
1/85	Exact subject and Exact corporate author searching introduced.
4/85	Inadequate disk-space leads to suspension of loading pre-1973 records
5/85	BROWSE and SELECT functions introduced
6/85	Ability to limit to last 10 years or last 3 years, or by language. Pre-1973 bloackade lifted.
8/85	TCP/IP installed to add computer power and flexibility.
10/85	Ability to show search history and redo search
1/86	2 million records in MELVYL catalog
4/86	EXPERIMENTAL MODE implemented: Experiment 1 changes CALLS display
6/86	MEDLINE Task Force meets.
7/86	Loading suspended due to floor-space and disk-space problems.
9/86	Ability to display browse results with citation counts.
10/86	Inhouse publications available online. Exact peridical title search. "Smart" title search.
12/86	50th issue of MOM.

2/87	ADDED SINCE curtailed
3/87	3 million records. MEDLINE test database, PARTIAL results for "long" searches. MOM online.
5/87	Ability to SAVE results to a list and DISPLAY LIST
7/87	NON-book materials loaded. Periodicaal results can be limited by campus
9/87	Subject searches getting 0 results prompted to try title-words
10/87	Automatic right truncation supplied for Exact subject heading searches. Find commands exceed 160,000 per week. Average search response exceeds 18 secs.
11/87	MEDLINE at selected terminals. MEDLINE Assist Mode introduced. Single letter truncation no longer accepted in periodical searches.
1/88	4 million records. Personal author algorithm refined to reduce response time.
4/88	FORM index allows searching and limiting by format of materials. Number indexes.
5/88	MEDLINE available in all UC libraries.
7/88	Specialized music indexes introduced. Lookup Mode changed to include "smart title search" and and prevent "long searches."
9/88	MEDLINE available remotely with passwords.
4/89	5 million records. Peak use restricitons forbid many common word searches.
6/89	USE CARL command gives MELVYL system users access to CARL databases.
10/89	READ, QUALIFY feature allows searches within some retrieved sets. More systems via USE.
3/90	MEDLINE expanded to 5 yrs.
4/90	CURRENT CONTENTS (CC) database loaded.
9/90	Separate enhanced periodicals database loaded, searchable by subject. TEN-year subset database.
10/90	Experiment 41: MEDLINE journals linked to library locations and holdings.
12/90	Ability to SAVE completed search results as SETS.
1/91	6 million records.
3/91	More USE systems, including EPIC. Find commands exceed 430,000 per week; average response less than 4 secs.
5/91	MAP indexes. Experiment 45: CC journals linked to library locations and holdings
6/91	Ability to limit by date for all years, including 1960+. Ability to search for Exact added titles.
9/91	3M Information Access Company databased added.

The MELVYL System and Its Academic Context

Stephen R. Salmon

One measure of the MELVYL system's success is the extent to which the system has achieved its academic goals. MELVYL, the system, was a key part of a plan adopted in 1977 to enhance the value of the University of California libraries to scholarship and research. The goal was to meld the libraries of the university's nine campuses, including the special, independent libraries as well as the nine general campus library systems, into a single library system—not administratively, but in the way the collections and services were offered to students, faculty, and staff.

If this goal could be met, the resources made available would be enormous—now numbering more than 25 million volumes. Clearly it would not be possible to use these resources as a single library, however, unless the users were able to see what was in the collections and where the desired items were located.

Toward this subsidiary goal of systemwide "bibliographic access," some previous steps had already been taken. A book catalog had been produced in 1972, but its forty-eight volumes contained only five years' cataloging and was out of date—by five years!—by the time it was published. Two campuses had converted their catalogs to COM (computer-output-microfilm), and initially it was assumed that the union catalog required to implement the new library plan would also be a COM catalog. In fact, three successive editions of a COM union catalog were produced, in 1979, 1980, and 1981, each one considerably larger than its predecessor.

To those of us who were worrying about the problem of bibliographic access (especially Mike Berger, whose creative ideas shaped much of the design of the MELVYL catalog, with the support and technical backing of Ed Brownrigg, Clifford Lynch, and many other members of the staff), a microform solution seemed inadequate, and we began to feel that the only practical solution was an online catalog. It hadn't been done before, and online systems in the library world were still relatively new, but our analysis convinced us that the approach was not only possible but realistic. In the end, the online catalog idea was embraced as the proposed means of access to the collections, and the proposal became a key recommendation in the plan adopted by the university and submitted to the California state legislature for funding.

The timing could not have been worse. The library plan called for millions of dollars in additional funding, not only for the online catalog but also for book purchases, buildings, and cooperative services. The request arrived on the legislature's doorsteps only days after the passage of Proposition 13, which cut property taxes drastically and confronted the legislature with monumental funding problems. To its credit, however, the legislature understood and appreciated the improvements in both service and efficiency that the plan, with its provisions for enhanced library cooperation, would make possible, and a few days later it granted all the funds requested.

The real test came a year or so later, when the campuses were forced to absorb the major cuts in their operating budgets that were caused, at least indirectly, by Proposition 13. The campus administrations passed these cuts on to the libraries as well as to other university departments, leaving it to the libraries to decide where best to take them. It was gratifying, if a bit surprising, that despite this pressure, no one suggested cutting the funds for the MELVYL catalog. The decision was partly economic—the online catalog had eliminated some significant costs, chiefly the filing of cards—but more importantly, the MELVYL catalog was beginning to meet the goal of providing bibliographic access to all the university's collections. This goal had been made more important and more urgent by the budget constraints. Book purchases on most campuses were likely to be slowed by budget cuts, for example, so it would be even more important to avoid unnecessary duplication in purchases and to rely more on cooperative collection development and cooperative use of the collections. For that approach to be realistic, the union catalog was essential.

In its early years, however, the success of the MELVYL catalog in meeting these academic goals was not a foregone conclusion, nor was the relation of the MELVYL project

Stephen R. Salmon is Chairman, Carlyle Systems, Inc., San Mateo, California.

to campus library operations and services always clear or agreed upon. For one thing, the technical approach was new, and there were plenty of skeptics convinced it would not work. For another, the project was a centralized one, operating out of the university's systemwide administration. In any such organization, there is always tension between the central headquarters and the "field" (in this case, campuses). Library operations had always been a campus matter, and the library staff's psychic rewards (to use a then-current phrase) came from offering the best library service possible and being appreciated for it. If some of that service came from a central source, that intimate campus relationship might be impaired, and the rewards diluted.

These fears sometimes produced surprising recommendations. At one point, for example, the Library Council (the campus library directors, the library school deans, and a few others) insisted that the union catalog should have access by main entry only. Other forms of searching, including subject access, should presumably be left to the campuses. Fortunately for all concerned, this issue became moot when the project managers reported that subject access in the online catalog had already been implemented. On another occasion, the campuses suggested that the system should be a distributed one, with satellite computers at each individual campus. In this case, at least in my view, the campuses were probably right.

The MELVYL catalog's role as a union catalog, and its place in the overall library scheme of things, was also blurred a bit by a suggestion in the plan that the MELVYL catalog might also serve in lieu of campus catalogs. The decision was left to the campuses, but the plan expected that the MELVYL catalog would replace all campus catalogs by 1984 or 1985. This happened at only three of the nine campuses. The MELVYL catalog simply was too complex, too big—and occasionally too slow—to serve this function well, and each campus now has its own online catalog system as well as the MELVYL union catalog.

Rather than attempt to serve as multiple campus catalogs, in fact, the MELVYL catalog has expanded, beneficially, in quite a different direction. It now includes, in addition to the holdings of the nine UC campuses, the California State Library, the California Academy of Sciences, the Center for Research Libraries, the National Library of Medicine's MEDLINE database, CURRENT CONTENTS, the Expanded Academic Index, the National Newspaper Index, the Computer Database, and the periodical holdings of the twenty-one campuses of the California State University, Stanford University, the University of Southern California, and the Getty Center for the History of Art and the Humanities. It also provides access to dozens of other library collections through an Internet gateway. The academic contribution of the MELVYL system—its support of study, scholarship, and research—has therefore expanded far beyond the original goal, and exceeds even the dreams of its original developers.

In only one respect has the academic influence of the MELVYL system been negative. MELVYL, the name, was a mild insider's joke, combining the spelling of WYLBUR, a Stanford utility used in the original system, with Melvil Dewey, who was not only the most famous figure in cataloging history but, ironically, a fierce advocate of spelling reform. Unfortunately, to judge from the number of requests for "Melvyl's Moby Dick" and other variations, a whole generation's spelling seems to have been corrupted. Perhaps this is a small price to pay, however, for the many academic benefits the MELVYL system has brought. ■■

Recent Publications

Book Reviews

Ardis, Susan B. *An Introduction to U.S. Patent Searching: The Process.* Englewood, Colo: Libraries Unlimited, 1991. 221p. paper, $32.50 U.S.; $39 outside North America (ISBN 0-87287-856-2).

At last, someone has presented the view that patent literature is not just for inventors. Susan Ardis argues, successfully, that the patent document is an important source of social and technical information, and demonstrates how the information contained in the millions of patent files could be used by librarians, social scientists, historians, archaeologists, and others interested in our technical past and future.

This volume is well organized; clearly written; and rich with examples, illustrations, definitions, further readings, and hints about searching techniques. Ardis begins with a general history of patents, examines the development of patents in the United States, discusses the essential elements of a patent document, outlines the structure of a patent office's files, and then devotes several chapters to the types of searches that a noninventor can perform and what kind of information would be retrieved from the searches. She also devotes a realistic chapter on the advantages and disadvantages of online patent searching. The book includes a bibliography, glossary, and a host of appendixes. Her discussion about the patent classification system is one of the best general analyses of this difficult subject that I have read to date. She is also thoughtful enough to describe briefly the Patent Depository Library system, and she gives a brief list of the participating institutions and their phone numbers.

What is most refreshing about *An Introduction to U.S. Patent Searching* is Ardis' ability to strip away the preconceived notion among nontechnical/science librarians and other researchers that searches of patent literature are limited to inventors or patent specialists. According to her introduction: "Historically, the primary use of patents has been to determine if something has already been patented. However, patents can also be used by historians, economists, sociologists, and others for such varied purposes as examining the development of specific areas of technological growth, researching the commercial interests of selected companies or the expertise and creativity of specific individuals, or for market research."

The book is highly recommended for all libraries.—*John A. Shuler, Colgate University, Hamilton, N.Y.* ■■

Brandt, D. Scott. *UNIX and Libraries.* Westport, Conn.: Meckler, 1991. 143p. $39.50 (ISBN 0-8876-541).

Technical works for systems librarians and information specialists abound but generally do not focus on practical applications for day-to-day library operations. This dearth has been most apparent for mainframe and minicomputer-based systems. Technical literature for personal computer applications is somewhat easier to come by, but as systems become more advanced, library staff are finding themselves managing increasingly larger and more complex system platforms. For those interested in a publication covering the UNIX operating system with a genuine library bent, this book is a welcome, albeit sparse, beginning.

As a tutorial, the book teaches the basics of UNIX shell and file structures. Brandt's focus is nontechnical, and experience in programming is not needed to take advantage of what his guide offers. Brandt is adept at illustrating uses of UNIX utilities from such a vantage point, especially when bridging basic system features to purposeful automated management of library data files. An adequate UNIX history is included, as are overviews of time-sharing and multitasking. There is no particular hardware platform focus.

Brandt's approach is to master the basics.

He covers the UNIX operating system, command files, database-management methods, electronic mail, and system security, all within a mere 130 pages. Within these pages, however, are enough tidbits to please the intended audience. Creative ideas include journal lists, bibliographies for new acquisitions, and interlibrary loan file maintenance.

Suggested approaches toward the building of online catalogs are a bit contrived, and dependence on the AWK and GREP utilities disregard current product offerings. The absence of references to electronic data transfer, the MARC format, and circulation solutions all serve to remind the reader that this book cannot be viewed as a tool for building an integrated system. Still, Brandt's solutions for text file management (many of which are not readily available from local system vendors) appear at once creative and easy to implement.

The brevity of the index will frustrate the novice, and the inclusion of a glossary might have been useful. The appendixes are valuable, as they bring together the names and addresses of vendors specializing in UNIX-based products for libraries and information centers. The exhaustive bibliography should serve as a fine reference for further collection development, although there is no inclusion of any of the currently available useful periodicals. Notable, too, is the absence of any material focusing exclusively on libraries, which indicates the value of Brandt's basic work.

While not likely to complement computer science collections, this book is recommended for all libraries and information centers utilizing UNIX solutions.—*David A. Jank, The New York Public Library.* ■■

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

When it became evident, some years ago, that libraries were finding it increasingly difficult to fill cataloging positions, many librarians set out to find out why and what could be done about it. The Committee on Education, Training, and Recruitment for Cataloging was formed within the American Library Association; several articles were written, and the Simmons College Symposium on Recruiting, Educating, and Training Cataloging Librarians was held in March of 1989. This is the second book to evolve from that symposium. The first publication, titled *Recruiting, Educating, and Training Cataloging Librarians: Solving the Problems* (Greenwood, 1989), contains papers prepared in advance of the symposium. This work consists of presentations or segments of presentations that did not appear in the participants' formal papers.

Keynote speaker Robert Hayes opened the symposium with some thought-provoking views. Although his comments were directed more toward librarianship in general, they succeeded in setting the tone for the symposium. He compared the profession of librarianship to that of medicine and proposed the following: that an undergraduate pre-library and information science curriculum be designed similar to pre-medicine and pre-law curricula; that a library school-teaching library partnership be developed much like the medical school-teaching hospital partnership; that the scientific foundations of librarianship be identified; and that the research base be expanded.

The remaining three parts are based on the arrangement of the symposium sessions: recruiting cataloging librarians, educating cataloging librarians, and training cataloging librarians. Each part includes statements of the session keynote speakers, observations of the other speakers, and comments that were made during the discussions following the formal presentations, which the editors organized into a number of theses around the topic. Included at the end of the book is an index, a selected bibliography, brief information about the contributors, and a list of audience members. As in most publications resulting from symposia, some of the presentations are more creative and better organized than others. In her presentation, Liz Bischoff calls for more "assertive recruiting" by employers, for example, talking to library school classes about the extent of opportunities that exist for catalogers and the kind of background needed. Recruiting techniques must be expanded and diversified, she contends. Others addressed the reorganization of libraries that could provide catalogers with responsibilities in public services and collection development. In this vein, James Neal spoke on the Penn State experience, which did succeed in making cataloging a more desirable choice at that institution. In areas where there are few

library schools to recruit from, distance education, which uses satellite communications and cable television networking, was put forth as a possible solution. Michael Carpenter proposed that a beginning cataloging course in library school teach what he termed "cataloging appreciation" to help students develop critical thinking, for example, how to analyze cataloging rules or evaluate cataloging systems. The technicalities of cataloging can come later. This opinion was supported in several later presentations.

This book should be useful to library managers and library educators. Indeed, if some of the recommendations are carried through, perhaps there will be no need for publications on this theme in the future!—*Ellen S. Kovacic, Hebrew Union College, Cincinnati, Ohio.* ■■

Information Technology: Design and Applications. Ed. by Nancy D. Lane and Margaret E. Chisholm. Boston: G. K. Hall, 1991. 354p. $39.95 (ISBN 0-8161-1908-2); paper, $24.95 (ISBN 0-8161-1909-0).

The editors state that the purpose of this book is to "describe the developments in information technology that are now becoming commonplace in business and industry, education and the arts," but with a focus "primarily on their present and future applications to the field of library and information science." The book is intended to "give students and practitioners sufficient understanding of these technologies to read more widely in the current journal literature."

Twelve articles are included, on data communications, networks, telecommunications applications, television and video, teletext and videotext, micrographic and optical disk technologies, CD-ROM and multimedia publishing, personal computer software, database management systems, artificial intelligence and expert systems, research on information access, and information policy and information technology. The editors inform us that they omitted an article on computers because "most students will be taking courses in introductory computing."

The editors' goals are met fairly well in this book. The articles are concise but provide good overviews and introductions to the topics, with citations to additional readings. While both students and practitioners must learn some technical terms when working with the technology, the authors of the articles do not overwhelm the reader with technical minutiae and jargon. The application of the technology in libraries and information centers is explained, although not in great detail. Some of the authors give the advantages and disadvantages of the technology in their articles. A comprehensive glossary defining technological acronyms and terms, including computer terminology, is provided.

There are some flaws, however, that might detract from the book's usefulness. A major problem is the dearth of diagrams or pictures illustrating the technology that is described in the text. There are a few illustrations, but most of the authors leave the visualization of the technology to the reader's imagination; good illustrations liberally used could have improved the text tremendously. It is puzzling why micrographics and optical disks are described in the same article. Also a puzzle is why the last two articles on research on information access and information policy and information technology were included. They are good articles, but somewhat out of context with the other ten; a good overview of computers would probably have better served the reader. The bibliographies of suggested readings could have been improved. Some of the twelve authors included short lists of allied readings, whereas others assume that readers will use their footnotes as sources for additional readings; none is comprehensive.

On the whole, this will be a useful book for the beginner, despite the absence of illustrations. It does provide overviews of various information technologies for the student and for the practitioner wanting to begin a basic background in information technology. It will not be useful to those already knowledgeable of the technology, except perhaps as a quick review. It will not give the reader a comprehensive introduction to the literature of the technologies covered, either.—*John Corbin, University of North Texas, Denton.* ■■

Johnson, Peggy. *Automation and Organizational Change in Libraries.* Boston: G. K. Hall, 1991. 201p. $39.95 (ISBN 0-8161-1919-8); paper, $24.95 (ISBN 0-8161-1920-1).

The purpose of this book is "to investigate the automation process and the impact of the computer on structure and environment in

academic libraries." It is an outgrowth of the author's twelve years of experience with a "seemingly endless stream of new technologies" at the University of Minnesota. Based on her belief that a paradigm shift is occurring in the organizational structure of libraries as a result of automation, this book documents a research effort to verify this perception. A survey questionnaire, which was sent to the heads of technical services at the member institutions of the Association of Research Libraries, was the basic data-collection mechanism for the research study. The survey questionnaire is included as an appendix.

The first two chapters give an introduction to the substance of the book by providing a brief history of the academic library and describing the current organizational environment. The focus is on the significant growth in academic libraries in the last forty years and their emphasis on collection building and user services. Academic libraries are described as service organizations in the larger university environment and, as a result, have been traditionally conservative, employing traditional hierarchical structures.

Chapter 3 describes the changes in library operations and services that are occurring as a result of automation. It focuses on the information explosion, increasing library costs, and the increased use of computer technology. Changes in both technical services and public services are described.

A detailed description of research results is given in chapter 4. Statistical tables are included that provide a straightforward statement of survey results and serve as the basis for the author's analysis. She finds that statistical evidence confirms the impact of automation on libraries, although few of these impacts are reflected in the formal organizational structure.

Organizational change is the focus of chapter 5. Literature on organizational change is cited to define change theory and the difficulty of implementing organizational change. Strategies are included for the introduction of planned technological change.

Drawing on data collected, chapter 6 describes how the responding libraries are implementing automation and, hence, organizational change. It is a detailed description of the process of purchasing and installing an automated system. Tables are used very effectively to show research results and to support the author's narrative.

The final chapters focus on the changes that are happening in libraries as a result of automation and suggest some techniques for moving into the future. There is strong support for libraries' use of the changing environment to take full advantage of the new opportunities being presented.

Automation and Organizational Change in Libraries is an excellent resource for any library with past, present, or future involvement with automation. It gives statistical evidence of the impact of automation on libraries, while describing the institutional impacts in terms of the latest theories of organizational change. For those planning for automation, it will give insight into impacts that are not described in the basic "how to's" of library automation. For those going through the automation process, it will help to provide a rationale for organizational changes. My primary concern with the book is that it focuses on technical services. A future edition, which evaluates automation impacts on public services and general library administration, would be a valuable addition to the literature.—*Pamela Q. J. André, National Agricultural Library.* ■■

Saettler, Paul. *The Evolution of American Educational Technology.* Englewood, Colo.: Libraries Unlimited, 1990. 570p. $38 (ISBN 0-87287-613-6).

This book is one of those uncommonly fine resources that will serve as an outstanding benchmark to anyone interested in the application and role of technology in education. Saettler, a preeminent educational technology historian, has done an excellent job extending and replacing the theoretical and historical dimensions of his 1968 work, *A History of Instructional Technology*, which, according to Ely's foreword, is a "prized possession" (p.xxv). Saettler's preface indicates that the book is primarily for professionals in the field and could also be used as a textbook. The purpose is to trace the theoretical and methodological foundations of American educational technology by examining the process rather than the product (devices and hardware) and also to provide an educational technology research overview as related to the public schools. Saettler is to be highly commended for his efforts in distilling a huge amount of information into something

extremely readable for his intended audience and purpose.

The book is organized into five parts, nineteen chapters, and two appendixes. The parts consist of the introduction (one chapter); "Heritage of Theoretical Thought and Practice, 450 B.C. to 1950" (seven chapters); "Growth of Theoretical Thought and Practice from Mid-Century to 1980"; "Educational Technology in the 1980s, 1990s, and Beyond" (four chapters); two appendixes, one on acronyms and the other a list of educational technology doctoral programs; and an index. One does tend to get caught up in the fascinating spin of Saettler's narrative in each chapter, and he does an excellent job of leading the reader through the era and process he is examining. Each of the chapters stands alone and encompasses the topics in great thoroughness. Readers or practitioners in the field would find this book a great resource even if they used only a fourth of what Saettler has written.

For me, the parts on the 1950s to 1980s and the 1980s, 1990s, and beyond were the most interesting. However, a minor criticism for this time period is lack of substantive reference to the tremendous implementation of educational technology by community colleges, where nearly half of all undergraduate students enroll. These colleges, as exemplified by Coast, Miami-Dade, Dallas, Portland, DuPage, Harper, Maricopa, Monroe, Brookdale, and others, pioneered effective learning resource centers, telecourses, technology/innovative consortia, and computing applications to instruction. Another concern is with chapter 14, where there is no recognition of library and information science's contributions to information science.

In his concluding chapter, Saettler leaves the reader wondering about the still-emerging theories and the fast-paced blends of newer technologies being rushed into practice, often without the thoughtfulness of past implementations. One also wonders who will be in and out of the field in the future. Perhaps, as Saettler says, "that things do not usually turn out the way they are predicted" (p.538) is the easiest prediction to make.—*George H. Voegel, William Rainey Harper College, Palatine, Illinois.* ■■

Schuyler, Michael, and Elliott Swanson. *The Systems Librarian Guide to Computers.* Supplement to Computers in Libraries, no.18. Westport, Conn.: Meckler, 1991. 130p. $39 (ISBN 0-88736-580-9).

This is an unusual book. The authors' objective was not to produce a reference guide, as the title would lead one to expect, but to write something that would console stressed systems librarians "with the knowledge that you are not alone" (p.viii, introduction). The result is a rambling narrative of the trials and tribulations of two librarians as they struggle to keep an online catalog and circulation system functional. (The authors are librarians at a regional library in Washington state.) Much of the material was previously published in *The Systems Librarian and Automation Review.*

Unfortunately, what seems like a good idea suffers from poor organization and a lack of judicious editing. The material has been assembled in a hodge-podge fashion that detracts from its readability and prevents the book from having any coherency. For instance, the first chapter—entitled "Living with the Beast"—lacks any introduction. Instead it opens immediately with a segment on how to compute the rip-off factor for memory boards, followed by three pages defending the use of a mouse, followed by a lamentation about how poorly vendors respond to RFPs. In another chapter, a blurb about how a community's activism stopped the construction of a microwave tower is inexplicably sandwiched between two segments on bar codes.

The authors' attempt to write in a breezy, informal way often results in text that is wordy and confusing, e.g., "the new issue also forced us to buy a copy of DOS 3 point whatever to accommodate the extension stuff" (p.90). Too many errors slipped by the editors, such as unnecessary hyphens (opposi-tionist [p.55]) and incomplete sentences ("In the first database management system I used to automate the membership lists of the Washington Library Association" [p.27].).

The first four chapters describe how the authors handled various automation-support situations. The text is liberally sprinkled with the authors' opinions on everything from consultants to equipment likes and dislikes. It is obvious that the authors learned their jobs the hard way—through trial and error—but that they have maintained their sense of humor. The last part of the book contains an interesting chapter entitled "Microsoft Book," a rather whimsical look at a microcomputer that

can be read and handled like a book. There also is an annotated bibliography on cyberpunk fiction and films that actually may encourage one to read some of this genre, although what cyberpunk has to do with the rest of the book is not explained. The way the bibliography is tacked on is indicative of the general disregard for organization that permeates the entire work.

This book will have limited appeal, especially at $39. It certainly cannot suffice as a reference work, and the predominance of acronyms, jargon, and technical terms (GODOS, ZOPL, dbms) without any explanations will limit its appeal to a general library audience. Systems librarians might enjoy reading some of the tales in this volume, but then they may have already seen them in *The Systems Librarian* and *Automation Review*. The authors were not well served by their publisher in this venture.—*Sally W. Kalin, Pennsylvania State University, State College.* ■■

Smith, G. Stevenson. *Managerial Accounting for Libraries and Other Not-for-Profit Organizations.* Chicago: American Library Assn., 1991. 212p. $50, $45 ALA members (ISBN 0-8389-0568-4) (ALA Order Code 0568-4-0010).

This is an excellent work describing different accounting methods for managing libraries and not-for-profit organizations. The author, G. Stevenson Smith, Ph.D., CPA, CMA, CCA, dedicates the work to "Everyone who derived benefit from my first book," *Accounting for Librarians and Other Not-for-Profit Managers* (American Library Assn., 1983). Since the conceptual basis on accounting for the two works (1983 and 1991) is the same, parts of the 1983 work are repeated in the work of 1991. However, in an expanded context, this work reflects the recent environmental changes in not-for-profit organizations.

As seen in the new title, the nature of this work has also changed from that of an introductory guide to a professional tool. The work introduces sound analytical tools for management accounting with a special emphasis on decision making. The reader familiar with the author's previous work will certainly have an easier entry to this work. On the other hand, this work alone will benefit anyone who is seeking professional assistance for improving his or her ability and skills for planning and making managerial decisions. The work is also extremely useful to the highest level of management of any not-for-profit organization in understanding the role of accounting in the organizational dynamics. The work is not a compilation of case studies grouped in certain categories nor a how-to book. It is a professional tool by which the reader will learn the conceptual bases and develop the ability to make critical analyses. The author does not assume the reader's knowledge or experience in accounting or fiscal control, nor does he demand strenuous effort to find useful information.

The selection of topics and their organization is analogous to a well-designed course. First, the author introduces conceptual bases, then he moves on to discuss standard and specific methods, and finally he proposes different applications and options. This progressive approach is particularly helpful to a reader who tends to look for a mere technique to justify decision making. The work helps the reader to do the job in reverse order, arriving at a decision by going through a logical process, from a broader context to a specific. The outcome of this process is by far more reliable. Included in the techniques and applications particularly pertinent to today's libraries are activities-based cost accounting, the flexible budget, developing life cycle management in the library, integrated library computer system, and the chapter on leasing.

The entire work is clearly written and readable. I strongly recommend the work for those who are engaged or interested in fiscal control of libraries or not-for-profit organizations, regardless of the size of the operation.—*Ryoko Toyama, Rutgers University, New Brunswick, New Jersey.* ■■

Virtual Reality: Theory, Practice, and Promise. Ed. by Sandra K. Helsel and Judith Paris Roth. Westport, Conn.: Meckler, 1991. 143p. paper, $39.95 (ISBN 0-88736-728-3).

This book, a reprint of articles published in the summer 1990 issue of *Multimedia Review* with several additions, defines *virtual reality* as "The creation of highly interactive, computer-based multimedia environments in which the user becomes a participant with the computer in a 'virtually real' environment" (back cover). Much like hypertext a few years

ago, virtual reality has become a trendy topic for discussion in much of the computer science literature.

Thus far there are few practical applications of virtual reality for the layperson or average organization. Existing applications tend to be very specialized, very limited, very expensive, or a combination thereof. The potential for the technology, however, is fascinating, as are the questions it raises. The eleven essays in this book, divided into "Theory," "Practice," and "Promise," while containing little immediately applicable information, are informative, challenging, and entertaining. They range widely from detailed discussions of issues to "gee-whiz" predictions of the future to philosophical discussions of the meaning of reality.

Michael Spring's essay in the "Theory" section, "Informating with Virtual Reality," is a good starting place for information specialists interested in the likely effect of these new technologies on their work. He focuses on virtual reality as a useful human-computer interface, discusses some of the major technical issues, and raises important practical questions for designing virtual reality interfaces for information storage and retrieval. While the essay is too short to serve as a full introduction, it includes a helpful bibliography for further study.

The more philosophic and predictive essays are, as a whole, less successful. With so much of the capability of virtual reality still potential rather than actual, the authors are left to philosophize or forecast with little basis in reality (however you choose to define it). This leads to interesting questions but shallow discussions. It also leads to considerable repetition among essays—the repeated mention of William Gibson's fictional view of cyberspace, for example. Those essays that take as their base technologies at least in the functioning prototype stage, such as the article on Fluxbase, are more successful.

The "Directory of Companies and Individuals" included is of little use. Not only are most of the addresses incomplete, but no explanation of each person's or organization's interest is given. In such a wide-ranging field, this is a critical omission. While brief, the "Recommended Readings" section provides a jumping off point for further explanation.

Despite these rough spots, the book succeeds in conveying the excitement and potential of virtual reality. Stewart Brand's *The Media Lab* would be a better introduction due to its greater thoroughness and consistency. As an adjunct to that, however, this book will serve admirably.—*Glenn P. Hoetker, SCAN C2C, Inc., Washington, D.C.* ■■

Software Reviews

DiscLit: American Authors. OCLC, 6565 Frantz Rd., Dublin, OH 43017; 1-800-848-5878, or G.K. Hall & Co., 70 Lincoln St., Boston, MA 02111; 1-800-343-2806. Hardware and software requirements: IBM PC, XT, AT, or PS/2 or compatible models, OCLC M3xx models, at least 640K RAM, DOS version 3.1 or higher, compact disc drive, and MS-DOS extensions 2.0 or higher. A black-and-white or color monitor may be used. Price: $995.

DiscLit on CD-ROM is a joint venture between OCLC and the publisher G.K. Hall & Co. Stored on a single CD, the product consists of two units: Twayne's United States Authors Series and the OCLC American Authors Catalog. The two units, referred to as the "Books" and the "Catalog," are searched separately, but the information in the two units is linked so that users can move easily between them. Published in book form by G.K. Hall, the Twayne's series is composed of 143 volumes, each on the life and works of a different American author, and each having that author's name as its title. The full text of the 143 books is contained on *DiscLit*. The other unit of DiscLit, the OCLC American Authors Catalog, contains bibliographic information for more than 100,000 books, serials, recordings, videos, and other materials by and about the 143 authors.

Installation of *DiscLit* is quick and easy, with clear instructions provided in the setup guide. An online tutorial is fairly well done, although to a first-time user of CD-ROMs it might seem overly complicated. The screen displays are quite cluttered. The tutorial's search tips are helpful, although some jargon terms, such as *limiters* and *navigation*, might be confusing to new users.

To search *DiscLit*, you may choose one of two searching modes: "new or casual user," or "experienced user." The first mode is a menu-driven system which is quite user-friendly. The user can enter a term or phrase, or can

combine terms by entering them on separate lines. For example, suppose a user wants information on American authors who have written on the topic of women and religion. Searching the full-text Books unit, the terms *women* and *religion* would be entered on separate lines, resulting in a list of authors' names with numbers indicating occurrences of the search terms. (Isaac Bashevis Singer was at the top with eighty-nine occurrences.) The system, by default, searches for the two terms in the same section of a chapter. What is a section? This is never defined on the CD or in the documentation. However, this default of combining terms by section has great potential for false hits. To avoid this, the user may specify that terms be in the same paragraph or within a certain number of words.

The search for women and religion resulted in a list of books, with those on Isaac Bashevis Singer, Sara Teasdale, and Emma Goldman having the most occurrences of the search terms. "Opening" the book on I. B. Singer, the user is first presented with the table of contents, with search word occurrences for each chapter. "Navigation" through the book itself is straightforward: chapters can be selected and opened, and moving through the chapters is facilitated by a very useful feature, "Jump to search word." This enables the user to skip irrelevant chunks of text and move to the next occurrence of the search word(s). Another handy way of moving around in a book is to go to the back-of-the-book index, select a page number for a particular topic, and jump to that page. One disadvantage of the mechanics of using the database: there is a heavy reliance on the Alt key with letter or function keys. For the experienced as well as the novice user, this seems cumbersome.

In the "Experienced User" mode, the user can create more sophisticated search strategies using the logical operators AND, OR and NOT as well as proximity operators and restrictors. Restrictors allow the user to limit a search to novelists, or poets, or other categories of authors. The proximity operators are especially useful in the full-text Books unit of *DiscLit*. The searcher can combine terms with authors' names, such as *rose* and *Gertrude Stein* to limit a search to occurrences of the term in only one author's work. The system is also useful if the user does not know an author's name or exact title. For example, the search *jumping para frog* looks for the *jumping* and *frog* in the same paragraph, and retrieves information on Mark Twain's short story, "The Celebrated Jumping Frog of Calaveras County."

While it takes a little while to find your way around the Books unit of *DiscLit*, the Catalog unit is very straightforward. The Catalog unit can be searched by term or author's name as well as by type of material (book, serial, AV media, etc.). Searching by type of material is a very useful feature of the system. For example, the user can enter *videocassette* and *Twain* to create a list of videos on Mark Twain and his works. Documentation for *DiscLit* is very good. There is a brief "Search Tips" card that can be set up next to the terminal; additional "Reference Cards" that summarize how to search, view, print, and save to a disk; and a detailed manual, *DiscLit User Guide*, which clearly explains all aspects of the CD-ROM. There is also an 800 number for an OCLC representative to answer questions.

High school and college students would certainly find *DiscLit* useful in doing research on American authors. The price of the CD is substantial, however, so its disadvantages must be considered. Most obvious is that not all American authors are included on the CD. Although 143 authors seems like a large number, the series attempts to include novelists, short story writers, poets, essayists, and playwrights, so it is certainly not comprehensive. Women authors are fairly well represented, but only seven African-American authors are included. This selectivity in the Books unit carries over to the Catalog unit, which contains bibliographic records not for all American authors as the title implies, but only for those 143 in the Twayne's series. Secondly, each book in the series has only one author, so students looking for a diversity of critical opinion will not find it here. Students often need several different sources of criticism for a particular novel, poem or short story; *Twentieth Century Literary Criticism* is one source that is invaluable for this purpose. On *DiscLit* users can find lists of critical sources, but only one source is presented in full text for each author. However, for the purpose of searching for themes, style, characters, and plot, *DiscLit* is very useful, as it is for the bibliographic information provided.—*Susan E. Clark, University of the Pacific.* ■■

World Atlas 2.0. The Software Toolworks, Inc., 60 Leveroni Court, Novato, CA 94949; (415) 883-3000. Hardware requirements: IBM PC or compatibles with a minimum of 640K RAM, hard disk (uses 6 megabytes), EGA or VGA graphics card with color monitor, Microsoft Mouse or compatible (supported but not required). Price: $59.95 (also available on CD-ROM, $79.95).

The Software Toolworks World Atlas user's guide says that "World Atlas is a powerful reference tool that combines state-of-the-art maps of every country in the world with an enormous database of international information that can be displayed and/or printed as text, maps, or graphs. It is an atlas, almanac, and world fact book in one easy-to-use program." The following review will attempt to evaluate how well the software lives up to its description.

INSTALLATION

The package reviewed provides eight 3.5-inch 720K diskettes for installation of the software. A toll-free number is given for those who need a 5.25-inch 1.2MB or 5.25-inch 360K diskette. Using a 3.5-inch high-density disk drive to load the 3.5-inch 720K diskettes presented no problems.

Installation proved to be simple following the step-by-step instructions in the user's guide. The INSTALL.EXE program displays a menu for the user during installation, giving exact instructions. The program also lists each file as it is being loaded with a real-time statistical display of the percentage of the file and program loaded. The display of this information is particularly helpful (especially to the novice) because it takes several minutes to load each disk.

The INSTALL program says not to worry if you err during installation because it is easy to start over. This is true ; after my first installation I got the "Too many open files" error message when I ran the program. Increasing buffers from 8 to 20 in my config.sys file solved the problem. The second installation went smoothly.

After installation, the user is told to check the READ.ME file that is created on the hard disk during the installation process. This file notes any last minutes changes. The user's guide instruction to type "Read.Me" to see this file does not work. What you actually get is a README.TXT file that you must type out to read, an annoying little problem for the novice. However, once you find the file, it does offer helpful suggestions as to how to use the PRINT function and how to increase the speed of the program.

The program cannot be run from a floppy disk drive, so a hard disk is required. It will use 6 megabytes of your hard disk.

OPERATION

The program begins by displaying a world map with a menu bar across the top of the screen (see figure 1). The menu bar does not include a HELP option. In fact, there are no help features in the program. From this, I assumed the program to be so user friendly that it would be self-explanatory. To test this theory, I followed the thirty-seven-page user's guide and learned all the features in a short period of time. The guide is well written and easy to follow. It includes many illustrations from the program.

Next, I chose a volunteer with an average amount of PC experience to use it without the assistance of the user's guide. He reported that he felt he had discovered all the program's abilities after about half an hour. I tested his knowledge and found that except for a few features, he was indeed able to use the atlas effectively without benefit of the user's guide. There are three program features that are not immediately obvious to the user. The use of the ESC key would be quickly discovered by a seasoned PC user, but not by the novice. When using World Atlas, the ESC key will close any window, take you to the closest index map, or back you out of the atlas. Another feature that you must discover for yourself (if the user's guide is not available) is using the mouse to click on the small square at the top left of windows to close the windows. ESC will also provide the same service. Also, when one of the menu bar options is not available at certain points in the program, the letters are faded, instead of simply not appearing on screen.

Certainly, it would be helpful if these three features were explained somewhere in the program. The missing information did not, however, seem to hinder my volunteer user. It appears from my small sample test that the program could be used satisfactorily by library patrons, most of whom would probably not access the user's guide.

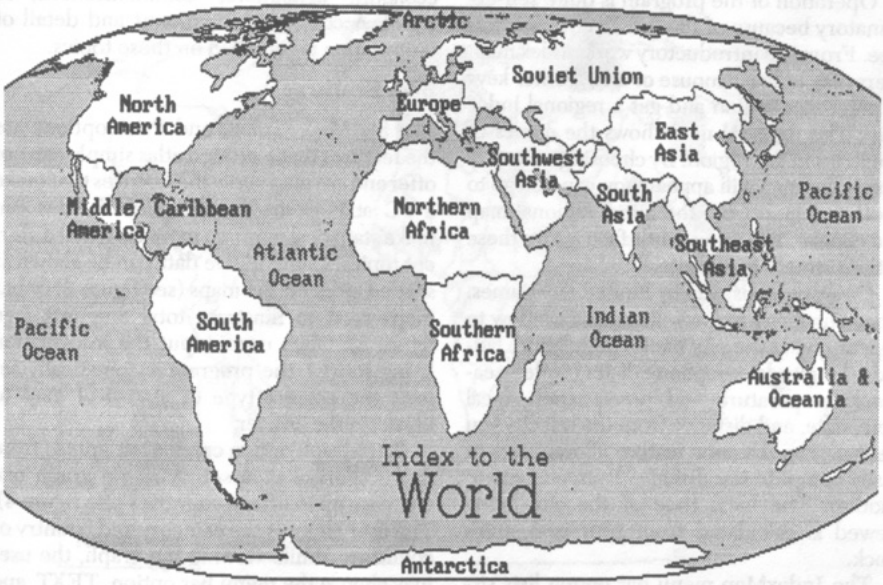


Figure 1. Initial World Atlas Display. Data are reproduced with permission from various United Nations and other international sources. A complete list of sources is included in the program and User Guide.

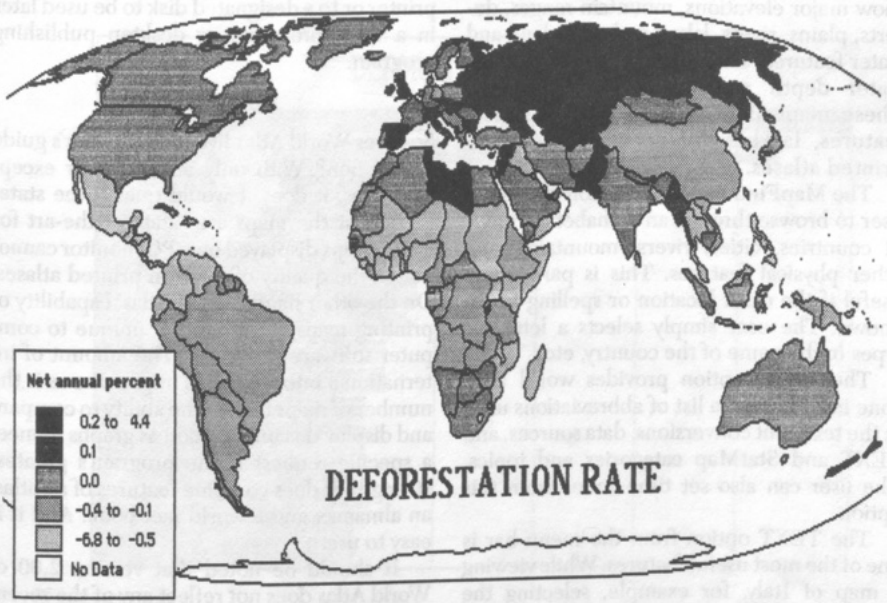


Figure 2. Statistical Map.

Operation of the program is quite self-explanatory because of the excellent user interface. From the introductory world index map, users may utilize a mouse or arrow/enter keys to point to a region and get a regional index map. The regional map shows the names of countries in the region. By choosing a country name, the map will appear. You can also go to another region from the initial regional map you choose. The mouse interface makes these actions simple to execute.

Country maps display limited city names. Choosing a city name will cause a window to appear giving the city name, population, latitude, longitude, telephone dialing codes, seasonal temperature and precipitation, local time, date, and distance from the last city you viewed. The distance feature allows users to easily compute the distance from one city to another. The local time of the city being viewed is calculated from your computer's clock.

The IndexMap menu bar option lists sixteen world regions from which you can select other maps. Ocean index maps are included. These maps show the oceans in the context of surrounding countries and regions.

The TopoMap option will allow you to look at eighteen relief maps for land regions that show major elevations, mountain ranges, deserts, plains, rivers, lakes, and other land and water features. Three ocean TopoMaps show water depth and ocean floor features. These maps show only major topological features, lacking the details expected in printed atlases.

The MapFind menu bar option allows the user to browse through an alphabetical index of countries, cities, rivers, mountains, and other physical features. This is particularly useful if the exact location or spelling is unknown. The user simply selects a letter or types in the name of the country, etc.

The Utility option provides world time zone information, a list of abbreviations used in the text, unit conversions, data sources, and TEXT and StatMap categories and topics. The user can also set time zones from this option.

The TEXT option from the menu bar is one of the most useful features. While viewing a map of Italy, for example, selecting the TEXT option will give you a window to choose information about the country's geography, people, education, health, government, crime, economy, agriculture, communications, or travel. A considerable amount and detail of information is provided on these topics.

Best Features

The StatMap, Graph, and Print options are the features that a printed atlas simply can not offer and are ultimately the features that make a PC atlas desirable. The World Atlas can make statistical maps on more than 300 different topics. Comparative data can be shown as shaded areas on the maps (see figure 2) or bar maps used to illustrate total amounts (see figure 3). Once users input the information being sought, the program automatically selects the correct type of statistical map to illustrate the answer.

The Graph option creates bar graphs from any numerical data set. A single graph can compare up to fifteen countries (see figure 4). The user chooses the category and country or countries. While viewing the graph, the user may choose the menu bar option, TEXT, and see the data from which the graph was made.

The Print menu option allows the user to print detailed reports that include maps, graphs, and text. Offering easy-to-follow printing instructions, the program provides the option of sending the map to a local printer or to a designated disk to be used later in a word processor or desktop publishing program.

SUMMARY

So, does World Atlas live up to its user's guide description? With only a few minor exceptions, yes, it does. I would qualify the statement that the maps are "state-of-the-art for PCs." Maps displayed on a PC monitor cannot match the quality of maps in printed atlases. On the other hand, World Atlas' capability of printing maps and graphs is unique to computer software programs. The amount of international information is impressive as is the number of maps (240). The ability to compare and display this information as graphs to meet a specific request is the program's greatest strength. It does combine features of an atlas, an almanac, and a world fact book. And it is easy to use.

It should be noted that version 2.00 of World Atlas does not reflect any of the recent changes that have occurred in the Soviet Union and eastern European nations. Software Toolworks representatives indicate that

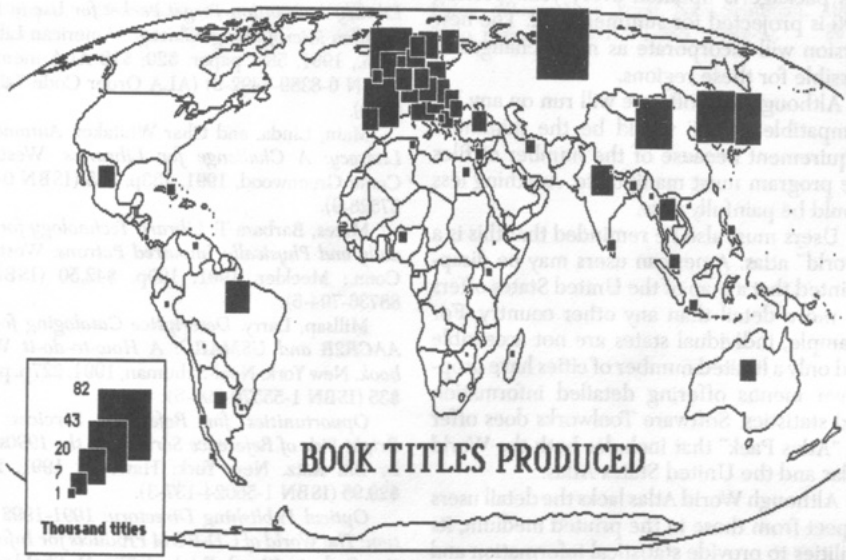


Figure 3. Bar Map.

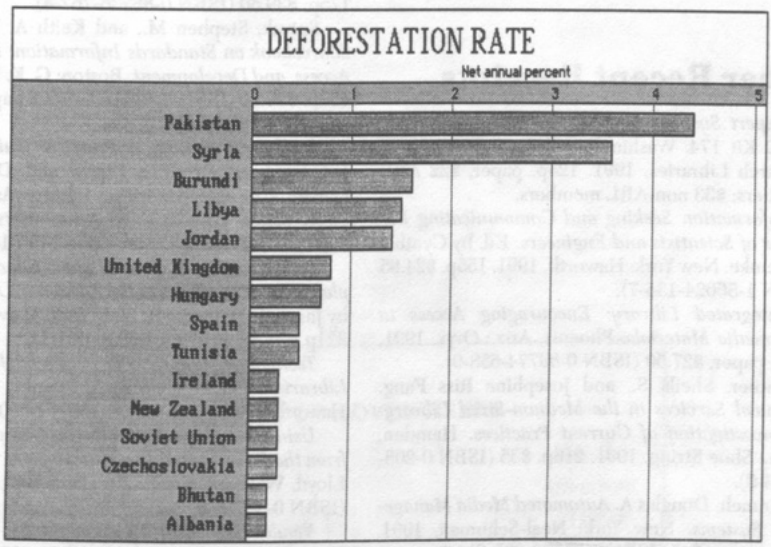


Figure 4. Bar Graph.

this package is updated every year. Version 3.00 is projected for summer 1992. The new version will incorporate as many changes as possible for these regions.

Although the software will run on any PC compatible, a 286 would be the minimum requirement because of the number of files the program must manipulate. Anything less would be painfully slow.

Users must also be reminded that this is a "world" atlas. American users may be disappointed that a map of the United States offers no more detail than any other country. For example, individual states are not accessible and only a limited number of cities have drop-down menus offering detailed information and statistics. Software Toolworks does offer an "Atlas Pack" that includes both the World Atlas and the United States Atlas.

Although World Atlas lacks the detail users expect from those in the printed medium, its abilities to provide statistical information and charts and to make excellent printed products specifically designed by and for the user make it a reference tool that greatly complements printed sources. Students and other PC users (especially geographiles) will find it extremely useful and fun.—*Lottie Simkins Meador.* ■■

Other Recent Receipts

Expert Systems in ARL Libraries, May 1991. SPEC Kit 174. Washington, D.C.: Association of Research Libraries, 1991. 125p. paper, $22 ARL members; $33 non-ARL members.

Information Seeking and Communicating Behavior of Scientists and Engineers. Ed. by Cynthia A. Steinke. New York: Haworth, 1991. 155p. $24.95 (ISBN 1-56024-135-7).

Integrated Library: Encouraging Access to Multimedia Materials. Phoenix, Ariz.: Oryx, 1991. 142p. paper, $27.50 (ISBN 0-89774-658-9).

Intner, Sheila S., and Josephine Riss Fang. *Technical Services in the Medium-Sized Library: An Investigation of Current Practices.* Hamden, Conn.: Shoe String, 1991. 216p. $35 (ISBN 0-208-02173-6).

Kranch, Douglas A. *Automated Media Management Systems.* New York: Neal-Schuman, 1991. 282p. paper, $45 (ISBN 1-55570-091-8).

Library Automation and Networking European Conference, May 1990. Ed. by The European Foundation for Library Cooperation. London: Bowker-Saur, 1991. 370p. $60 (ISBN 3-598-10935-0).

McKern, Debra, and Sherry Byrne. *American Library Association Target Packet for Use in Preservation Microfilming.* Chicago: American Library Assn., 1991. 58p. paper, $20; $18 ALA members (ISBN 0-8389-7492-9) (ALA Order Code 7492-9-0010).

Main, Linda, and Char Whitaker. *Automating Literacy: A Challenge for Libraries.* Westport, Conn: Greenwood, 1991. 133p. $35 (ISBN 0-313-27528-9).

Mates, Barbara T. *Library Technology for Visually and Physically Impaired Patrons.* Westport, Conn.: Meckler, 1991. 190p. $42.50 (ISBN 0-88736-704-6).

Millsap, Larry. *Descriptive Cataloging for the AACR2R and USMARC: A How-to-do-it Workbook.* New York: Neal-Schuman, 1991. 227p. paper, $35 (ISBN 1-55570-098-5).

Opportunities for Reference Services: The Bright Side of Reference Services in the 1990s. Ed. by Bill Katz. New York: Haworth, 1991. 213p. $29.95 (ISBN 1-56024-137-3).

Optical Publishing Directory: 1991-1992 Edition: The World of CD-ROM Products for Information Seekers. 4th ed. Ed. by James H. Sheldon and Joseph A. Webb. Medford, N.J.: Learned Information, 1991. 293p. paper, $59 (ISBN 0-938734-54-7).

Saffady, William. *Optical Storage Technology 1990-1991: A State of the Art Review.* Westport, Conn.: Meckler, 1990. 230p. $49.50 (ISBN 0-88736-594-9).

Search Sheets for OPACs on the Internet. Ed. by Marcia Henry. Westport, Conn.: Meckler, 1991. 175p. $39.50 (ISBN 0-88736-767-4).

Spivak, Stephen M., and Keith A. Winsell. *A Sourcebook on Standards Information: Education, Access, and Development.* Boston: G. K. Hall, 1991. 451p. $39.95 (ISBN 0-8161-1949-X); paper, $27.95 (ISBN 0-8161-1948-X).

Staff Development: A Practical Guide. 2d ed. Ed. by Anne Grodzins Lipow and Deborah A. Carver. Chicago: American Library Assn., 1991. 104p. paper, $25; 22.50 ALA members (ISBN 0-8389-3402-1) (ALA Order Code 3402-1-0010-0).

Strategic Planning in Higher Education: Implementing New Roles in the Academic Library. Ed. by James F. Williams II. New York: Haworth, 1991. 221p. $24.95 (ISBN 1-56024-091-1).

Technology Transfer: The Role of the Sci-Tech Librarian. Ed. by Cynthia A. Steinke. New York: Haworth, 1991. 172p. $22.95 (ISBN 1-56024-116-0).

Using Computer Networks on Campus: Papers from the First Annual Conference 1990. Ed. by Les Lloyd. Westport, Conn.: Meckler, 1991. 137p. $30 (ISBN 0-88736-813-1).

Vendors and Library Acquisitions. Ed. by Bill Katz. New York: Haworth, 1991. 235p. $29.95 (ISBN 1-56024-121-7).

The Whole Library Handbook: Current Data, Professional Advice, and Curiosa about Libraries and Library Services. Comp. by George M.

Eberhart. Chicago: American Library Assn., 1991. 500p. paper, $25; $22.50 ALA members (ISBN 0-8389-0573-0) (ALA Order Code 0573-0-0008).

Yerkey, A. Neil. *Information Management Using dBase.* New York: Neal-Schuman, 1991. 206p. paper, $35 (ISBN 1-55570-094-2). ■■

INDEX TO ADVERTISERS

Be a PR Star! enter the

1993 JOHN COTTON DANA LIBRARY PUBLIC RELATIONS AWARDS CONTEST

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

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

The John Cotton Dana Award

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

The Special Award

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

Contest Dates

Entries for the 1993 John Cotton Dana Library Public Relations Awards Contest can reflect any one of the following time frames:

- Calendar year 1992 (January-December)
- School Year 1991/92 (Fall-Spring)
- Special Project which ends in 1992.

The Deadline for entries is February 1, 1993.

Awards Ceremony

Official award citations will be presented to contest winners at the 1993 annual conference of the American Library Association, at a reception hosted by The H.W. Wilson Company.

Sponsorship

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

To Enter

To request an Information Packet containing contest entry forms, rules and regulations, questions and answers about the awards, a sample of the judges' evaluation form, names of the contest judges, and a list of previous winners, please write to: John Cotton Dana PR Awards Contest, The H.W. Wilson Company, 950 University Avenue, Bronx, New York 10452.

Letters

To the Editor:

In her review of my book *Indexing and Abstracting in Theory and Practice* (*ITAL*, December 1991, p.343–44), Bella Weinberg took pains to point out that, of my 331 bibliographic references, 26% date from the 1960s and 32% from the 1970s. She also complains that the bibliography is incomplete. One possible interpretation is that I am unaware of the uncited sources. To set the record straight: as she well knows, substantial bibliographies of the literature on indexing (containing several thousand entries) already exist. It is not the responsibility of the writer of a textbook to compile a complete bibliography, but merely to include references to items that he/she considers particularly important, or items that illustrate some point the author wishes to make. It is a poor textbook that indiscriminately includes references to everything. If I fail to cite an item it is more likely to mean that I consider it not worth citing (perhaps because it says nothing new) than that I am unaware of it. In my opinion, the only really significant literature on subject indexing to be published in the last 30 or so years comprises the reports of the Cranfield studies and Coates' *Subject Catalogues: Headings and Structure* (1960). Robert Fugmann has developed some useful axioms of indexing, but they deal more with the vocabularies used in indexing than they do with the act of indexing. A legitimate case could be made for the claim that nothing truly original in indexing has been published since Ranganathan's work in the l930s. One might even claim that no really major contribution to the literature has been made since Cutter.—*F. W. Lancaster, University of Illinois.*

To the Editor:

In the December 1991 issue of *Information Technology and Libraries* (v.10, no.4) Yem Fong reviewed *The Journal of Interlibrary Loan & Information Supply.* Prior to Dr. Fong's review, I was unsympathetic to athletes, movie stars, and politicians who were always whining about being "quoted out of context." Now I understand their complaint.

Dr. Fong quoted me as writing, "the editor states that, 'he always objects to the emergence of a new journal,' " as a reason for Dr. Fong's objecting to the founding of *The Journal of Interlibrary Loan & Information Supply.*

What I actually said was, "Like you, the Editor always objects to the emergence of a new journal. Enough already. But there are new fields, new topics, new concerns, and new emphases. Interlibrary loan librarians deserve their own journal. Interlibrary loan is not a subfield of reference services, public services, circulation, or any other field." Ironic humor is sometimes difficult to grasp.

Dr. Fong states that *The Journal of Interlibrary Loan & Information Supply* "overlaps to some degree with *Interlending & Document Supply.*" The editors of *Interlending & Document Supply* choose all of the topics, and then choose authors to write on those topics. *The Journal of Interlibrary Loan & Information Supply* gets all of its material "over the transom." In other words, the interlibrary loan librarians decide, individually, who will write and about what they will write.—*Leslie R. Morris.*

To the Editor:

I beg to differ with one of Walt Crawford's conclusions in "Tutorial: Desktop Publishing Choices: Making an Appropriate Decision" (Dec. 1991, *ITAL*). I appreciate most of the advice offered but suggest that his experiences using word-processing software in 1988 or earlier may not be relevant to today's programs. In fact, my experiences have been the opposite of his. I used Ventura Publisher for a number of years (it seemed like decades) from its first release until forced to choose between the Windows and the Gem versions (ca. 1990). In that time, I used it for a number of professional papers, annual reports and brochures, and database publishing (e.g., *Directory of Microcomputer Interests among Ohio Librarians* [Ohio Library Association, 1988]). In those days, you didn't

have to buy a special Ventura DataBase Publisher at $199.

I have since switched to WordPerfect (WP) 5.1 using Bitstream FaceLift fonts (far cheaper than the PostScript capability Crawford recommends and, in its most recent version, offering tolerable fonts even on a 24-pin dot-matrix printer). I once again am able to concentrate on content and let WP take care of most of the formatting. In the past few months, in addition to the types of materials for which I previously used Ventura, I have used this combination for my and my wife's *Historical Dictionary of Singapore* (Scarecrow Press, 1991) and Hwa-Wei Lee's collection, *Librarianship in World Perspective: Selected Writings, 1963–1989* (Taipei, Taiwan: Student Book Company, 1991). Both were done with camera-ready copy (one "real" size and one reduced by the publisher) and included scanned images and computer-generated indexes. My one regret was using an H-P Laserjet Series II rather than the Laserjet III family with Resolution Enhancement Technology to smooth jagged lines and curves.

I would agree that Ventura, PageMaker, and the other desktop publishing (dtp) programs mentioned may be useful for newsletters, especially those with cutesy graphics; however, for more serious textual material, I will stick to powerful word processors, such as WP 5.1. The learning curve is slashed: special capacities are learned incrementally as needed. Importantly, text is massaged within the program in which it was created rather than needing to be imported into a dtp program. This greatly facilitates editing and revising, although I hear that Ventura has greatly improved its text-handling capabilities.

While I support my preference in terms of capabilities and efficiency in usage, cost cannot be ignored. Many libraries can qualify for Word Perfect's educational discount at $125 or $130 (a fraction of that through its School Purchase Program), and by taking advantage of offers, a good collection of FaceLift (or other scalable fonts) costs less than $100. *Chacun à son gout*, but for productivity, ease-of-use, and cost-effectiveness, I will stick with a diversified word processing capability rather than an over-specialized dtp dinosaur.—*K. Mulliner, Assistant to the Dean, Ohio University Libraries, Athens.* ■■

SEE YOU IN DENVER!

Information Technology: IT's for Everyone

Library and Information Technology Association
Third National Conference
Denver, Colorado
September 13–16, 1992

More than 50 sessions

Designed for all types of librarians and information professionals:

- American Memory Project
- Client/server architecture
- Computer viruses
- Electronic networking
- Ethics and privacy
- Libraries as publishers
- Multiple language systems
- Virtual reality

Research Forum • Showcases • New Product Reviews

❒ YES! Send me a copy of the Preliminary Program and Registration Information for the LITA Third National Conference!

Name _____

Address _____

_____

City _____

State _____ Zip _____

❒ Please send me membership information.
❒ Please send me information on LITA publications.

Return this form to:
LITA Third National Conference 1992
50 E. Huron Street
Chicago, IL 60611-2729
FAX: (312) 280-3257

SETTING STANDARDS IS EVERYONE'S BUSINESS.

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

Which is why we need your help.

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

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

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

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

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

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

NISO

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