Cumulating the Supplements to the Seventh Edition of LC Subject Headings

Roy B. TORKINGTON: Central Library and Documentation Branch, International Labour Office.* At the time of writing, the author was head, Library Systems Department, University of California, San Diego.

A description is presented of the project of the University of California Library Automation Program to cumulate the 1966 through 1971 supplements to the Library of Congress Subject Headings. The University of California Institute of Library Research MARC processing software, BIBCON, was used, with specially written programs. The resulting cumulation was edited, printed in book form, and made available to libraries. The final task involved merging six MARC files into one file of over 125,000 records and then printing that file in a format similar to that of LC Subject Headings. The project was a cooperative effort with participation by people from several UC campuses.

INTRODUCTION

The seventh edition of *Subject Headings Used in the Dictionary Catalogs of the Library of Congress* was published with a cutoff date of June 30, 1964. The first supplement covered additions and changes from July 1, 1964 through December 31, 1965. Subsequent supplements were issued annually, with each annual cumulating quarterly over a one-year period.[1,2] By 1972, when the supplement for 1971 was issued, it was necessary, when assigning or verifying subject headings, to use seven supplements (1964/ 65, 1966, 1967, 1968, 1969, 1970, 1971) in addition to the seventh edition. The Supplement Cumulation Task of the University of California Library Automation Program aimed at alleviating that problem by merging the 1966 through 1971 annual supplements into one cumulation. Through the courtesy of the Library of Congress we were able to obtain unedited magnetic tape files of all supplements except the 1964/65; these files were in the LC internal format.

The University of California Library Automation Program undertakes cooperative automation programs which will benefit UC libraries. The cu-

| SUPCON | SKED | SORT | SUPMRG | SUPER BIBLIST BIBPRINT | SUPUNK | Manual Edit | FIX | SKED | SORT |

Printed 66-70 Preliminary Cumulation

FIX Statements

66-70 LC Annuals — 66-70 UC — Keyed 66-70 UC Annuals — Sorted Keyed 66-70 UC Annuals — Keyed Unedited 66-70 Cumulation — Unedited 66-70 Cumulation — Edited 66-70 Cumulation — Keyed Edited 66-70 Cumulation — Sorted Keyed Edited 66-70 Cumulation
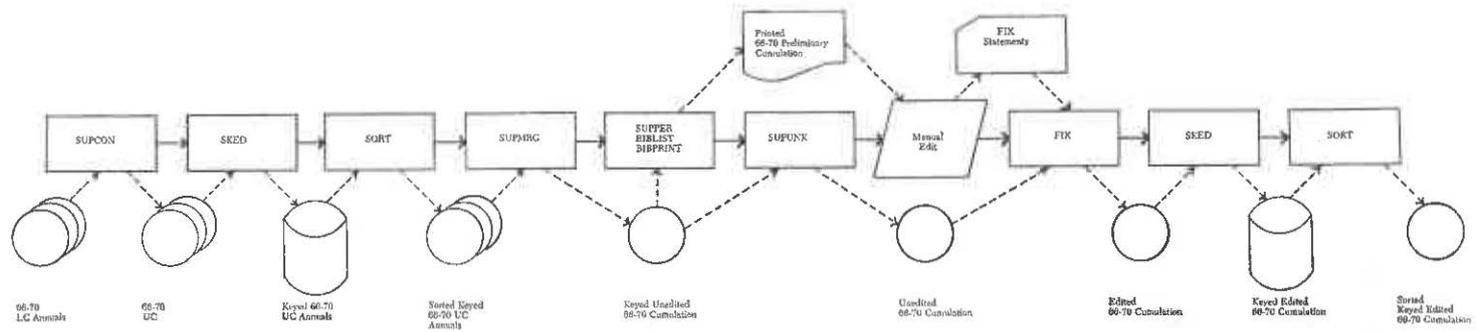
Fig. 1. *Preliminary cumulation production and editing.*

| BIBLGEN | SKED | SORT | SUPER BIBLIST BIBPRINT | Manual Edit | BBL | SKED | SORT | SUPMRG | SUPER BIBLIST BIBPRINT |

UC Annual

Old Keyed Cumulation

LC Annual — Keyed UC Annual — Sorted Keyed Annual — Printed Annual — FIX Statements — Edited Annual — Keyed Edited Annual / Sorted Keyed Edited Annual — New Keyed Cumulation — New Printed Cumulation
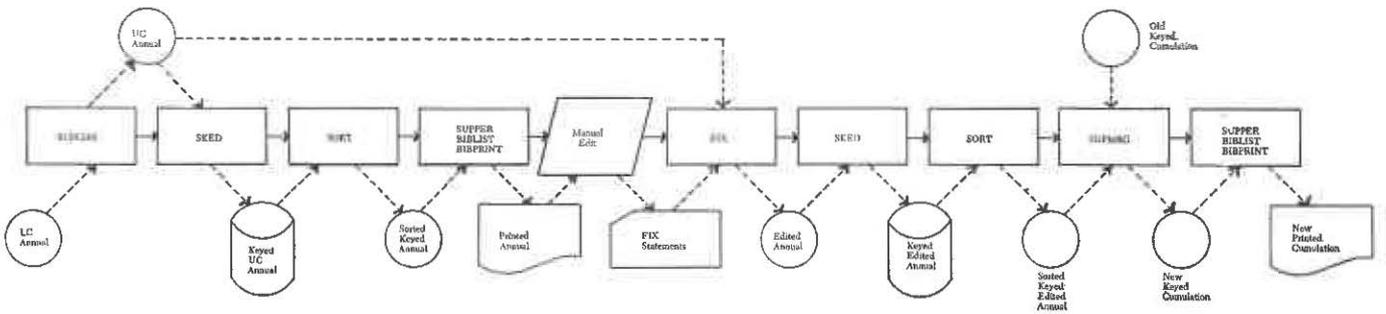
Fig. 2. *Annual supplement cumulation cycle.*

mulative supplement task was seen as a low-cost project which could result in a product useful to UC and other libraries. We intended to use software already available at UC as much as possible. The available software was BIBCON, a group of MARC processing programs developed and used by the UC Institute of Library Research, Berkeley, and UC Santa Cruz, for production of the 1963–1967 *UC Union Catalog Supplement* and the UC Santa Cruz book catalog.[8] The BIBCON programs to be used were: SKED, a sort key creator and editor; BIBLIST and BIBPRINT, record, column, and page formatters; and FIX, a record corrector. Essentially we considered the problem of cumulating the supplements as one of producing a book catalog by cumulating several annual catalogs. Thus we needed a program to convert the LC internal format to UC MARC (SUPCON) and a merge program (SUPMRG) (see Figures 1 and 2).

MERGE TRANSACTIONS

A special merge program was necessary, because merging the supplement files is more complex than merely interfiling entries. Frequent matches or "hits" modify each other. In fact, the merge logic was critical to the whole process and dictated the design of the record format. Assume that we have the following sequence of headings given in the seventh edition of *Subject Headings* and modified as shown by the 1966, 1968, and 1970 annuals supplements. (The word "Number" refers to classification numbers associated with the headings.)

Seventh Edition	Heading (Old Number) xx Tracing D

1966 Supplement	Heading
	    xx Tracing A
	         Tracing C

1968 Supplement	Heading (Direct) (New Number)
	CANCEL
	Heading (Old Number)
	    xx Tracing B
	CANCEL
	    xx Tracing C

1970 Supplement	CANCEL
	Heading (Direct) (New Number)
	CANCEL
	    xx Tracing A
	CANCEL
	    xx Tracing B
	CANCEL
	    xx Tracing D

We planned the merge to take place annual by annual, beginning by merging the 1967 supplement into the 1966, producing a 1966–67 cumulation. The 1968 supplement was then merged into the 1966–67 cumulation and so on, proceeding until all available files were cumulated. With the example given above, the following steps would be used to produce the final cumulation:

1966–67 Cumulation	No change from 1966 annual

1966–68 Cumulation	Heading (Direct) (New Number)
	CANCEL
	Heading (Old Number)
	  xx Tracing A
	     Tracing B

1966–69 Cumulation	No change from 1966–68
1966–70 Cumulation	CANCEL Heading (Old Number) CANCEL xx Tracing D

After the final cumulation we see that the heading and its class number and tracing have all been canceled. A user checking both the seventh edition of *Subject Headings* and the cumulated supplement would see that the heading and its tracing had been canceled and would not use it.

The examples given demonstrate many of the transactions which occur in cumulating the supplement. Class numbers and tracings are added and deleted, direct/indirect status is changed, headings are canceled. Other transactions involve adding new headings and subdivisions, reinstating previously canceled headings, and canceling headings added in previous supplements. In merging supplements into an old master file (i.e., producing an eighth edition), one would delete from the file a canceled heading or tracing. However, when cumulating a supplement one must differentiate between a cancellation and a deletion. Thus, in the example, the 1968 supplement canceled Tracing C. Since Tracing C was added by the 1966 supplement, the net result for the cumulation is to delete Tracing C (i.e., remove it from the file). In the 1970 supplement, Tracing D is also canceled. Since this tracing was added by the seventh edition the cancel information for it must remain in the file. A cancel transaction in a supplement can thus result in either deleting data from the cumulated file or adding data. If the original data first appeared in the seventh edition and is later canceled, the cancel record remains in the cumulated supplement. If the original data first appeared in a supplement and is canceled in a later supplement, the record is deleted, i.e., removed from the cumulative supplement. If the cumulative supplement were to be merged into the seventh edition file, all cancels would be deletes.

RECORD FORMAT

At first we gave serious consideration to defining the record so that one heading (or heading and its subdivisions) and all its references and tracings occurred as separate fields and/or subfields in one MARC-structured record. This creates a serious problem in the merge when adding and deleting references and tracings. Each of the transactions shown in the example would require adding data to, or removing data from, the variable fields of a MARC record. We decided to simplify the record structure along the lines of LC's internal format and thus simplify the merge program.[4]

Since any heading, subdivision, reference, tracing, note, or class number in a supplement can be added, canceled, or deleted, the format is defined so that every one of those elements would establish a record. Direct/Indirect information is carried in the heading or subdivision record as one character in the fixed field. Also, each record for a subdivision, class number, reference, tracing, or note contains the headings and subdivisions under which it files. Each element occurs as a variable field in the MARC record. Consider the following example from the 1967 supplement:

> Building materials
>   sa Structural engineering
>   —Research (Direct) (TA404.2)
>   xx Materials research

This is represented by five MARC records with variable fields tagged as shown in the following example. (Examples are only representations of MARC records. Tags actually occur in the directory. Indicators and delimiters associated with variable fields are not shown.)

Record 1:	150 Building Materials
Record 2:	150 Building Materials
	710 Structural Engineering
Record 3:	150 Building Materials
	201 Research
	(Direct Indicated in Fixed Fields)
Record 4:	150 Building Materials
	201 Research
	050 (TA404.2)
Record 5:	150 Building Materials
	201 Research
	711 Materials Research

Cases in the printed supplements where a heading is canceled and then reestablished in order to change class number or Direct/Indirect indication required special definition. Thus:

> Heading (Direct) (Number)
> CANCEL

	Heading (Number)
Became one record:	150 Heading
	(Direct Indicated in Fixed Fields)
Also:	Heading (New Number)
	CANCEL
	Heading (Old Number)
Became three records:	150 Heading

150 Heading
050 (New Number)

150 Heading
050 (Old Number)
(Cancel Indicated in Leader)

Table 1. UC MARC Subject Supplement Variable Fields

001	LC locator number	
008	Fixed fields	
045	Special sorting field	
050	LC classification number	
150	Heading	
201	Subdivision 1	
202	Subdivision 2	
203	Subdivision 3	
204	Subdivision 4	
500	Scope Note	
501	Example or note under	
510	General see note	
520	General sa note	
700	See reference	
701	x tracing	
710	sa reference	
711	xx tracing	
790	First See reference	Temporary tags set
791	First x tracing	and used only by
792	First sa reference	SUPPER and BIBLIST
793	First xx tracing	programs
999	Special tag reserved for SKED and BIBLIST programs	

MERGE LOGIC

The UC SKED program can generate 256-byte sort keys that will cause each record to sort in the desired sequence. Because of program and fund limitations, certain types of filing were ignored, e.g., chronological, geographic, diacritics, and some hyphenated headings. For acronyms, initialisms, and some abbreviations, we inserted special filing fields as part of the postmerge edit process. Generally, however, we were able to come quite close to LC *Subject Headings* filing by strictly programmatic means. Since the merge program logic requires records in sequential order by key, it was necessary to sort the files after adding keys to them, to correct for the difference between key order and LC *Subject Headings* filing.

The greatest variation from LC *Subject Headings* filing practice occurs in our treatment of chronological and geographic subdivision. Instead of filing in chronological order:

U.S.
  —History
  —Revolution
  —1783–1865
  —Civil War

we file in alphabetical order:

U.S.
U.S.—History
U.S.—History—1783–1865
U.S.—History—Civil War
U.S.—History—Revolution

Surprisingly, there have been very few complaints from librarians about this filing method. It does not appear to have impaired the utility of the cumulation.

The merge program (SUPMRG) is shown in Figure 3. Annuals are merged into previous cumulations until an annual record (ANNREC) and a cumulation record (CUMREC) have equal keys (AKEY=CKEY). For records with very long fields, usually scope notes, the key may require more than 256 characters. In these cases it is truncated. In only one case out of a series of merges involving over 100,000 records were truncated keys matched.

As can be seen from Figure 3, the record format makes it possible for SUPMRG to merge the files without modifying any variable fields. SUPMRG, as a result, does not require modules for rebuilding and modifying MARC records.

Figure 3 shows that a match of reference, tracing, etc., records is error free only when one record is a cancel and the other is not. Otherwise the system is attempting to duplicate an add or a cancel. However, in the case of headings and subdivisions it is possible, and indeed common, to have matches of records which are not cancels. This occurs whenever new tracings or references are added to a heading or subdivision already in the cumulated file. In our first example the first records from the 1966 supplement will be: 150 Heading. From the 1968 supplement:

150 Heading
(Direct Indicated in Fixed Fields)

SUPMRG sets the Direct/Indirect indicator, writes a record to output, and continues merging. An add-cancel match with references and tracings results in no record in the cumulation (e.g., Tracing C in the same example); the two records delete each other. In the case of headings and subdivisions, however, we cannot be certain that a canceled heading was added in a supplement and is not in the seventh edition. Hence, headings and subdivisions are never deleted. This leaves a few ghosts which should be deleted, i.e. canceled headings which were added in supplements and later
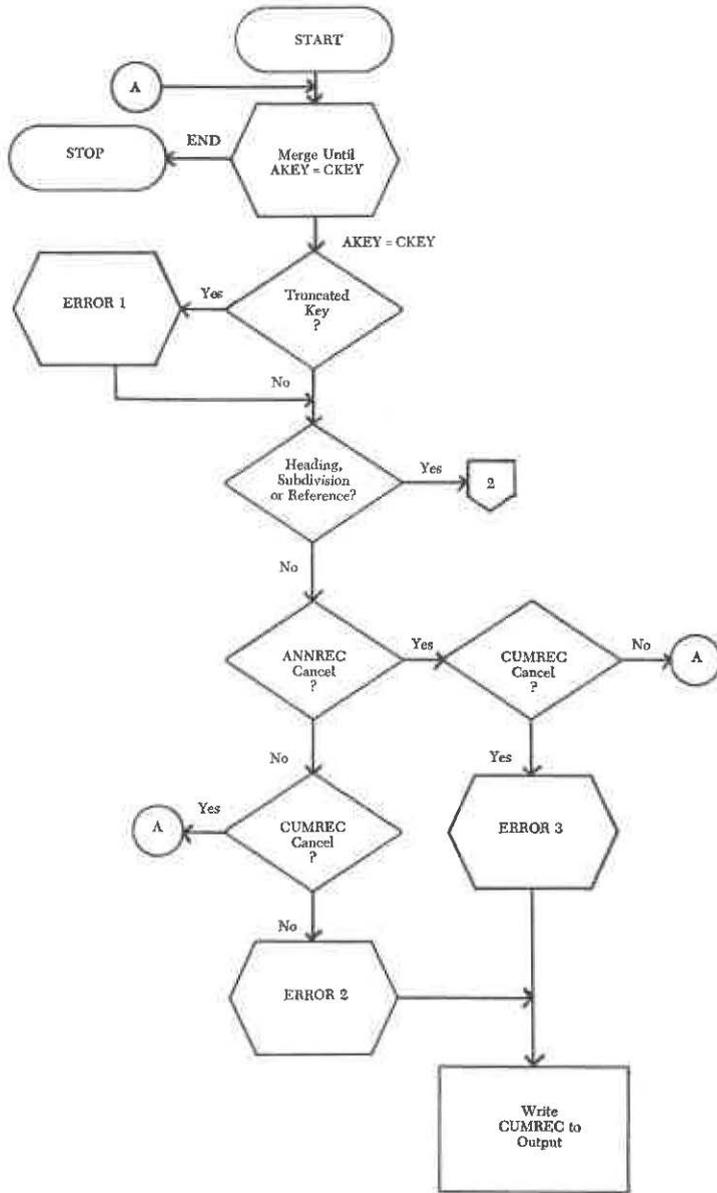
Fig. 3. Supplement merge program (SUPMRG).

canceled. The problem was judged to be unavoidable and relatively harmless.

PRINT FORMAT

Example 1 shows a sample of printed output from the cumulation. It varies from the printed format of LC *Subject Headings* as shown in Ex-
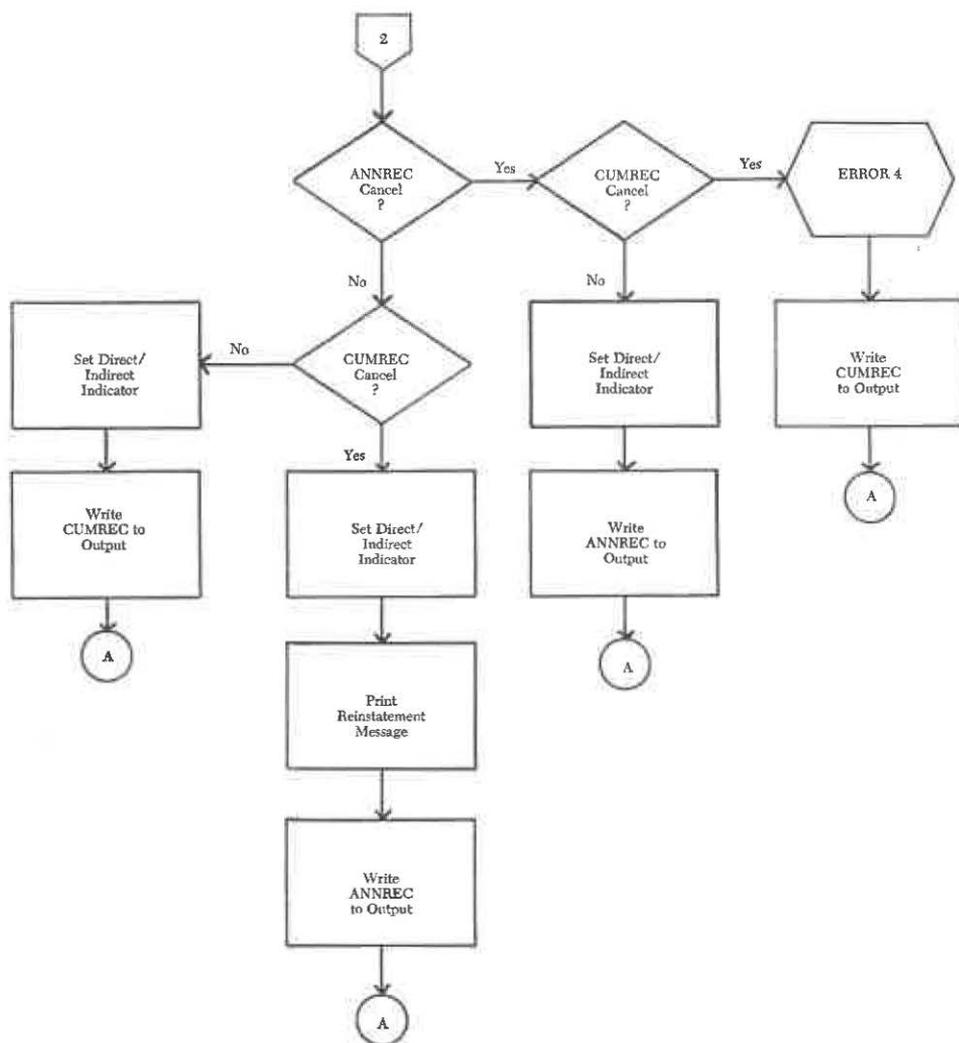
Fig. 3 (continued)

ample 2. In the cumulation, headings and subdivisions are repeated, only one type font is used, and diacritics are ignored both in printing and filing. Also, placement of (Direct), (Indirect) and class numbers differs from the LC style.

The cumulation was formatted and printed using three programs: SUPPER, BIBLIST, and BIBPRINT. BIBLIST and BIBPRINT are part of the Institute of Library Research (ILR) BIBCON software package and were designed for printing book catalogs. SUPPER was written specially for the supplement cumulation to prepare records for BIBLIST; it places records containing variable fields for (Direct), (Indirect), and

```
Botany, Economic
    (Indirect)
CANCEL
        Works on the economic botany of
            individual localities are entered under
            1. Botany, Economic. 2. Botany-- local
            subdivision
    sa  Fiber plants
        Fungi--Economic aspects
CANCEL
    sa  Textile fibers
Botany, Experimental
    x   Experimental botany
Botany, Medical
Botany, Medical--Pictorial works
Botflies
    sa  Sheep botfly
Botswana
Botswana--History
Botswana--History--1966-
Botswana--History--To 1966
Bottle gardens
    See Glass gardens
Bottle tickets
    See Wine labels
Bottlenosed dolphins
        (QL737.C4)
    xx  Dolphins
Bottles
    (Direct)
        (Art, NK5440; Technology, TP866)
CANCEL
        (TP866)
    sa  Bitters bottles
        Miniature bottles
        Poison bottles
    x   Flasks
    xx  Glassware
Bottles--Collectors and collecting
    (Direct)
Bottles--Collectors and collecting--Caricatures
        and cartoons
```

Example 1

Bosnia and Herzegovina †
　—History †
　　—1878-1918
　　—1918-

Boston Tea Party, 1773 †
　—Juvenile literature

Botany †
　—Classification †
　　sa Plants—Identification *
　—Ecology †
　　sa Island flora
　—Experiments
　—Identification *
　　See Plants—Identification
　—Variation †
　　sa Island flora

Botany, Economic †
　sa Fungi—Economic aspects

Botswana *
　—History
　　—To 1966
　　—1966-

Bottle tickets *
　See Wine labels

Bottles *(Direct)* *(Art, NK5440;*
　　Technology, TP866) †
CANCEL:
Bottles *(TP866)* †
　xx Glassware

Bottles, American, [etc.]
　x American [etc.] bottles

Bottom relief, Ocean
　See Submarine topography

Boulou language *
　See Bulu language

Boundary layer control †
　sa Vortex generators *

Boxing accidents
　See Boxing—Accidents and injuries

Boys †
　sa Fathers and sons *
　　Mothers and sons *
　—Prayer-books and devotions
　　(BV283.B7) †
CANCEL:
　—Prayer-books and devotions †
CANCEL:
　　Subdivided by language.
　　x Boys—Prayer-books and
　　　devotions—English
　—English
　　See Boys—Prayer-books and
　　　devotions
　—French, [German, Italian, etc.]

Braid †
　sa Braiding machinery

Braiding machinery *(TS1782.5)*
　xx Braid
　　Textile machinery

Brain †
　sa Corpus callosum
　　Globus pallidus
　　Subthalamus
　—Abnormities and deformities †
CANCEL:
　　Example under Deformities
　—Blood-vessels †
　　—Surgery
　—Diseases †
　　sa Cerebral edema
　—Edema
　　See Cerebral edema
　—Electromechanical analogies
　　xx Brain—Models
　—Models †
　　sa Brain—Electromechanical
　　　analogies
　—Research
　　See Brain research
　—Surgery †
　　sa Frontal lobotomy

Example 2

CANCEL in appropriate places in the merged file, and also strips keys from records.

SKED and BIBLIST were originally designed to interface in producing book catalogs. In processing a MARC bibliographic master record, one wishes to explode the master record so that there is a keyed record for every subject heading and added entry. The records are then sorted on key

and printed. With bibliographic data the first element of the key will be the first element printed. Title records are filed on title, and the title is the first printed element of a title entry; similarly, with name and subject entries. The key itself is so edited by SKED (all uppercase, no punctuation, superblanks added) as to be useless for printing. As a result, sorted records coming into BIBLIST must have an indication of the tag which generated the first element of the key. This is accomplished by use of a 999 tag in the MARC record directory which contains the tag and sequence number of the appropriate field. SKED adds the 999 tag to the record and BIBLIST uses it to format print records.

The problem with using this interface in printing the cumulated supplements is that all keys begin with the 150 (i.e., heading) field, whereas print lines for tracings, references, etc., do not. The solution was to have SKED add the 999 tags, and have SUPPER modify them so that they point to the tags for the fields which are to be printed.

Another problem with BIBLIST occurred when a series of, say, xx tracings occurred under a heading. We wanted to follow LC *Subject Headings* format and print an "xx" preceding only the first "see also from" tracing in the column. BIBLIST cannot distinguish between the first record containing a 711 variable field (i.e., xx tracing) and following records with 711 fields. Either all would have xx printed or none would. To solve this problem we set up a special set of tag values for first x, and xx tracings, and first See and sa references. These values were used only by the print program. SUPPER changed the appropriate directory entries, and BIBLIST used them in formatting lines.

Obviously BIBLIST could have been modified to do all this processing. However, we judged it easier and quicker to write a relatively simple key stripper and tag changer (i.e., SUPPER) than to modify a complex, table-driven program such as BIBLIST.

COORDINATION

The UC Library Automation Program (LAP) is a cooperative effort involving representatives from all nine UC campuses and a central program office. The people working on the cumulation were scattered from San Diego to Berkeley. Three different IBM 360 computers were used, two of them under the OS, one under the DOS operating system.

This decentralization was not the result of poor planning, but was done deliberately. The cumulation project was a minor task of LAP and there was not staff available at any one location to work on it. The only solution was to farm tasks out to available people, of whom there were few. Since the project manager was at San Diego and did not have access to an IBM 360 running under OS, his programs (SUPCON, SUPMRG, SUPPER, SUPUNK) were written to run on an available DOS system. Eventually, if LC does not cumulate supplements to the eighth edition of *Subject Headings*, the programs will be converted to run under OS. No program-

mer has been available to do this yet. Likewise, it was impossible to install the BIBCON programs at San Diego, since they ran under OS. Hence SKED, BIBLIST, and BIBPRINT were set up and run elsewhere.

Initially, this project was conceived as part of a test of cooperative development. In this case, cooperation was successful. However, a task of systems development in line operations, such as acquisitions or circulation, involving large scale changes in library procedures, might not be successful under such conditions. This project was a relatively small but interesting task which did not involve changes in library operations.

Running on more than one computer and under two different operating systems caused problems. Tapes were continually sent by intercampus bus between Santa Barbara and San Diego, a distance of 200 miles. The problem was compounded when multireel files were involved and we wished to label them. Unfortunately, IBM 360 DOS tape labels and IBM 360 OS tape labels are mutually incompatible—a noteworthy nuisance. Program testing was also difficult. Frequently, the best test of a program is to run its output through the next program in the system. When the next program is 200 miles away and a tape must be mailed, delays become routine.

CUMULATION SYSTEM

Figures 1 and 2 depict the cumulation system. Figure 1 shows the steps involved in producing a preliminary 1966–70 cumulation. This cumulation was edited by part-time student assistants using procedures developed by a librarian working on the project. Corrections were made by the BIBCON program FIX, and the corrected cumulation was rekeyed and sorted. Figure 2 shows the annual cycle. We began this cycle with the 1971 supplement and the edited 1966–70 cumulation. A new cumulation (in this case, 1966–71) was printed. The new cumulation is retained with sort keys and becomes next year's old cumulation, obviating the necessity of rekeying and resorting this large file (over 125,000 records).

COSTS

The cost of the project was less than six thousand dollars, including salary costs of editors and keypunchers, but not of the project manager at UCSD and the programmer at UCSB who worked on this project along with several others, contributing varying amounts of time. The cost includes the cost of all computer runs, from testing and debugging through production. Printshop costs are excluded. However, the cost of computer runs to print out reproducible copy for both the 1966–70 and 1966–71 cumulations is included. The cost of one annual cycle (excluding printshop and labor costs) is about five hundred dollars. It should be noted that the cost of a quarter-time analyst and programmer will easily double these costs.

CONCLUSION

The cumulation was well received by librarians; copies have been sold, to non-UC libraries, grossing more than enough to repay the cost of the project. Future plans are extremely dependent upon LC's cumulation pattern after the eighth edition of *Subject Headings*. If LC cumulates its supplements annually or if it brings out new editions at frequent intervals, say, every three years, there will be no more need for the UC Cumulation and it will be discontinued.

REFERENCES

1. U.S. Library of Congress. Subject Cataloging Division, *Subject Headings Used in the Dictionary Catalogs of the Library of Congress* (7th ed., Washington: Library of Congress, 1966) and its *Supplements.*
2. Robert R. Holmes, "Introduction to the Seventh Edition of Subject Headings Used in the Dictionary Catalogs of the Library of Congress," *Library Resources and Technical Services* 12:323–29 (Summer 1968).
3. Liz Gibson, "BIBCON; A General-Purpose Software System for MARC-Based Book Catalog Production," *Journal of Library Automation* 6:237-56 (December, 1973).
4. Henriette D. Avram, Lenore S. Maruyama, and John C. Rather, "Automation Activities in the Processing Department of the Library of Congress," *Library Resources and Technical Services* 16:195–239 (Spring 1972).