

DOCUMENT RESUME

ED 118 070

IR 002 971

AUTHOR Bourne, Charles P.
 TITLE Planning Data for the Conversion of UCUCS-2 Catalog Card Records Into Machine Readable form (UCUCS-2 Conversion Planning). Report No. ILR-75-006.
 INSTITUTION California Univ., Berkeley. Inst. of Library Research.
 REPORT NO ILR-75-006
 PUB DATE Jan 76
 NOTE 41p.

EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage
 DESCRIPTORS *Cataloging; Catalogs; Data Bases; Libraries; Library Automation; Library Planning; *Library Research; Library Technical Processes; *Union Catalogs; *University Libraries
 IDENTIFIERS Machine Readable Cataloging; University of California

ABSTRACT

The Institute of Library Research converted to machine-readable form over one million catalog cards representing about 750,000 Roman-language titles cataloged by the University of California (UC) libraries from 1963 to 1967. From this data base they printed the UC Union Catalog Supplement (UCUCS-1). From 1968 to 1972, an additional set of card records was collected and preprocessed. This study determined the extent to which the new UCUCS-2 records are available in already existing University of California data bases and in a few "outside" data bases. A one-half percent stratified sample was drawn from each UC campus and findings revealed that 82% of the UCUCS-2 sample could be associated with some kind of unique identification number, but only 63% of the total actually had this number on the catalog record; the others had to be searched in other files. (Author/CH)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. Nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

ED118070

ILR-75-006

PLANNING DATA FOR THE CONVERSION OF UCUCS-2 CATALOG
CARD RECORDS INTO MACHINE READABLE FORM
(UCUCS-2 CONVERSION PLANNING)

Charles P. Bourne

Institute of Library Research
University of California
Berkeley, California

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

January 1976

2

R 002 971

CONTENTS

ABSTRACT

FIGURES

TABLES

I. INTRODUCTION

II. OBJECTIVES

III. METHOD OF APPROACH

A. Development of UCUCS-2 Sample

B. Search for Overlap in Existing Machine Files

1. Search in UC Files

2. Search in Ohio College Library Center (OCLC) File

IV. FINDINGS

A. Total Source Records to be Converted

B. Nature of the Source Records to be Converted

1. Language

2. Imprint Date

3. Availability of Unique Identification Number

C. Availability of Source Records in Existing Machine Files

1. Availability in the UC Files (LC MARC, UCSC, and UCUCS-1)

2. Availability in the OCLC File

3. Availability in Other Files

ABSTRACT

In 1970 the Institute of Library Research converted to machine-readable form over 1 million catalog card records representing about 750,000 unique Roman-language titles cataloged by UC libraries during the period 1963-1967, and from this data base printed the UC Union Catalog Supplement (UCUCS-1). Some 1.7 million additional Roman-language card records representing monographs cataloged by UC libraries between 1968-1972 (here called UCUCS-2) have been collected and manually pre-processed by ILR. Equivalent machine language records now need to be obtained for this new group of records. This study determined the extent to which UCUCS-2 records are available in already existing UC data bases and in a few "outside" data bases.

A 1/2% stratified sample was drawn from each UC campus, totalling 8,337 records. (Records at UC Santa Cruz are already in machine-readable form and so were not considered in this sample.) Of these, about 48% could be found in one of the machine files already available at, and used by, the UC system: about 27% were available in LC MARC files, about 9% in the University of California Santa Cruz (UCSC) files, and about 12% could be found in the UCUCS-1 files. Of those not found in any of the 3 UC files, about 30% (an estimated 217,000 records) could be located in the OCLC data base. These findings indicate that of the total 1,697,822 UCUCS-2 Roman-language records, at least 807,000 records could be copied from existing machine files (these figures being adjusted to allow for 9% of the source records found to be out of the defined sample scope --e.g., non-monographic materials).

The sample was also analysed by language, by imprint date, and by availability of a unique identification number, such as ISBN or LC Card Number. 81% of the UCUCS-2 sample could be associated with some kind of unique identification number, but only 63% of the total, or about 1.06 million records, actually had this number on the catalog record, the others having to be searched in other files.

FIGURES

<u>Figure</u>	<u>Page</u>
1. Sample Record with Data Recording Template	5
2. Flowchart of Record Search Process for UC Machine Files	8
3. Age Distribution of UCUCS-2 Monograph Records	18
4. UC File Overlap for Riverside UCUCS-2 Records	28
5. UC File Overlap for Davis UCUCS-2 Records	29
6. UC File Overlap for Irvine UCUCS-2 Records	30

TABLES

<u>Table</u>	<u>Page</u>
1. Distribution of Publication Dates in Full LC MARC Data Bases as of July 1975	9
2. Materials Received for UCUCS-2	15
3. Distribution of English and Non-English UCUCS-2 Monograph Records	17
4. Age Distribution of UCUCS-2 Monograph Records	19
5. Proportion of UCUCS-2 Records with LC Card, ISBN, or ISSN Numbers	21
6. Number of Records with a Unique National or International Identification Number	22
7. Summary of Machine Record Availability from External Data Bases for UCUCS-2 Records	25
8. UCUCS-2 Monograph Records Identified as LC MARC Records	26
9. Number of Residue Records Included in OCLC Data Base	32

I. INTRODUCTION

The Institute of Library Research (ILR) has already converted to machine readable form over 1 million catalog card records representing about 750,000 unique Roman-language monographic titles cataloged by UC libraries during the period 1963-67. From this computer data base, the 47-volume UC Union Catalog Supplement (here referred to as UCUCS-1) was produced and is now in use in each of the nine UC campuses, the California State library, and each of the 19 California State University and Colleges System campuses.

Some 1.7 million additional Roman-language card records, representing monographs cataloged by the UC libraries during the period 1968-72 have also been collected and pre-processed by ILR, and the cards are now warehoused in the Richmond storage facility awaiting further planning and processing. These records are referred to as UCUCS-2 records. (All of the UC Santa Cruz records are already on computer tape and will be available when appropriate for further UCUCS-2 processing.)

There seems to be agreement within the UC Library System that the UCUCS-2 records now warehoused in Richmond should be converted into machine readable form. The question is how to convert these records as economically and easily as possible. Since no firm commitment has yet been made to any specific product or service (e.g., printed or microform author-title book catalog, or on-line searching), this proposed machine file ought to have the flexibility to generate a wide variety of products or services such as catalog cards, circulation records, a systemwide shelf list, and printed book catalogs or supplements.

Some of the card records are already available in machine form from external source files such as the LC MARC, OCLC, UCUCS-1, and UC Santa Cruz machine files, and could very likely be copied therefrom with less time and cost than required for original conversion. Some records are not available in machine form and would have to be converted.

Ken Weeks, who headed the project at its start, was responsible for most of the data collection effort for this study, and was later assisted by other graduate students of the UCB School of Librarianship, including Nancy Christenson, Ron Heckart, Ned Himmel, Jane Irby, and Bob Treppa..

II. OBJECTIVES

The major objectives of this study were to:

1. Determine the magnitude of the UCUCS-2 conversion problem
2. Determine the extent to which UCUCS-2 records were already available on some existing data bases

Additional lesser objectives were to:

1. Determine the extent to which the UCUCS-2 records overlapped with several existing machine bibliographic files
2. Determine the nature of the library materials represented by the UCUCS-2 catalog records (i.e., the nature of the material acquired and cataloged by the UC campuses during the period 1967 through 1972)
3. Determine the extent to which some unique record identification number such as the LC Card Number was available for each of the UCUCS-2 records.

III. METHOD OF APPROACH

A. DEVELOPMENT OF UCUCS-2 SAMPLE

All of the UCUCS-2 records from the UC Santa Cruz campus library already exist in machine readable form as a byproduct of that library's continuing local book catalog production efforts, and do not pose any significant conversion problem for the subsequent UC union catalog or union file efforts. Consequently this study was really interested only in the catalog records from the other eight campuses.

In order to provide an accurate and meaningful basis for our analysis efforts, a 1% stratified random sample was drawn from the UCUCS-2 cards submitted by each of the campuses other than Santa Cruz. Because all of the UCUCS-2 cards which had previously been sent to ILR had gone through some preliminary processing (e.g., re-packaging, numbering, microfilming) we had relatively good estimates of the total number of catalog cards submitted by each campus. This helped the sampling plan considerably. Because we had a good estimate of the total number of cards from each campus, we could then easily specify a sample size to be drawn from each campus' UCUCS-2 records. And because all of the cards had been stamped with a unique number by a numbering machine it was relatively easy to use a random number table to draw numbered records from each campus' records in order to meet the specified sample size. The selected records were extracted from their containers and copied onto a template such as that shown in Figure 1. A total of 8,337 records were copied in this manner to be used as the source data for this study.

B. SEARCH FOR OVERLAP IN EXISTING MACHINE FILES

1. Search in UC Files

There are three existing UC machine files that can help with the UCUCS-2 conversion effort. All of these data bases are immediately accessible to the UC University-wide Library Automation Program (ULAP) with no extra acquisition or royalty costs: the full retrospective LC MARC data base is already stored at the Data Processing Center in Berkeley and used on a regular production basis by Bibcenter; the UCSC data base can be obtained readily from UCSC when needed; and the UCUCS-1 data base presently resides at ULAP and plans are being made to prepare it for regular Bibcenter production searching.

It was assumed that in the event that a record might be available from more than one machine file, there would be a preference for one file over the others. It was agreed that the most desirable file (because of its high quality and detail of tagging of bibliographic elements and its ready availability at low cost to UC) would be the LC MARC file. This is clearly the first choice among the available files. The next choice (because of its relatively clean and complete MARC format records) is the UC Santa Cruz machine data base covering the entire UCSC library holdings. The third choice would be the UCUCS-1 data base, with its lesser quality and level of tagging,

Because of this preferential selection policy, most of our searching of the sample records was done on a sequential basis, following the stated preferences. A record was searched initially to see if it was in the LC MARC data base; if it was, the search terminated at that point. If the record was not in the LC MARC data base, it was then searched in the UCSC data base, and so on. A few of the sample records were searched in all of the data bases during an early part of this study in order to estimate the relative overlap in the several machine data bases.

The search process usually followed the sequence illustrated in Figure 2. The first step was to examine the sample record to see if it was really a catalog record for monographic material. The original UCUCS effort was intended by policy to include only monograph materials. However, a significant amount of non-monograph material actually was submitted by the campuses and included in the printed book catalog. The exact scope of the UCUCS-2 conversion effort has not been established yet. However, for the purpose of this study we made the simplifying assumption that only monograph records would be included. With this assumption we then pulled from the sample all non-monograph catalog records as well as all other records that had slipped past the initial UCUCS-2 pre-processing activities (e.g., catalog records with non-Roman characters, notes, and order slips). This excluded material turned out to be a significant fraction of the sample.

The remaining records were immediately examined to see if by inspection they could be positively identified as being in the LC MARC data base. This is indicated by the notation "LC MARC" or "MARC" on the card. These cards were immediately set aside in a separate pile.

The remaining cards were then examined to see if they were "probable MARC" cards. Because of the way in which retrospective and continuing conversion efforts are being made at the Library of Congress (LC) there is no publication date at which one can absolutely rule out the possibility of an LC MARC record. However, there is some frequency distribution information for publication dates that can provide some useful guidance. Table 1, for example, obtained from a recent unpublished analysis of the complete LC MARC file, shows the number of LC MARC records as of mid-1975 that had the indicated publication dates. Few records are available with publication dates before 1960, and LC was generating very few foreign language catalog

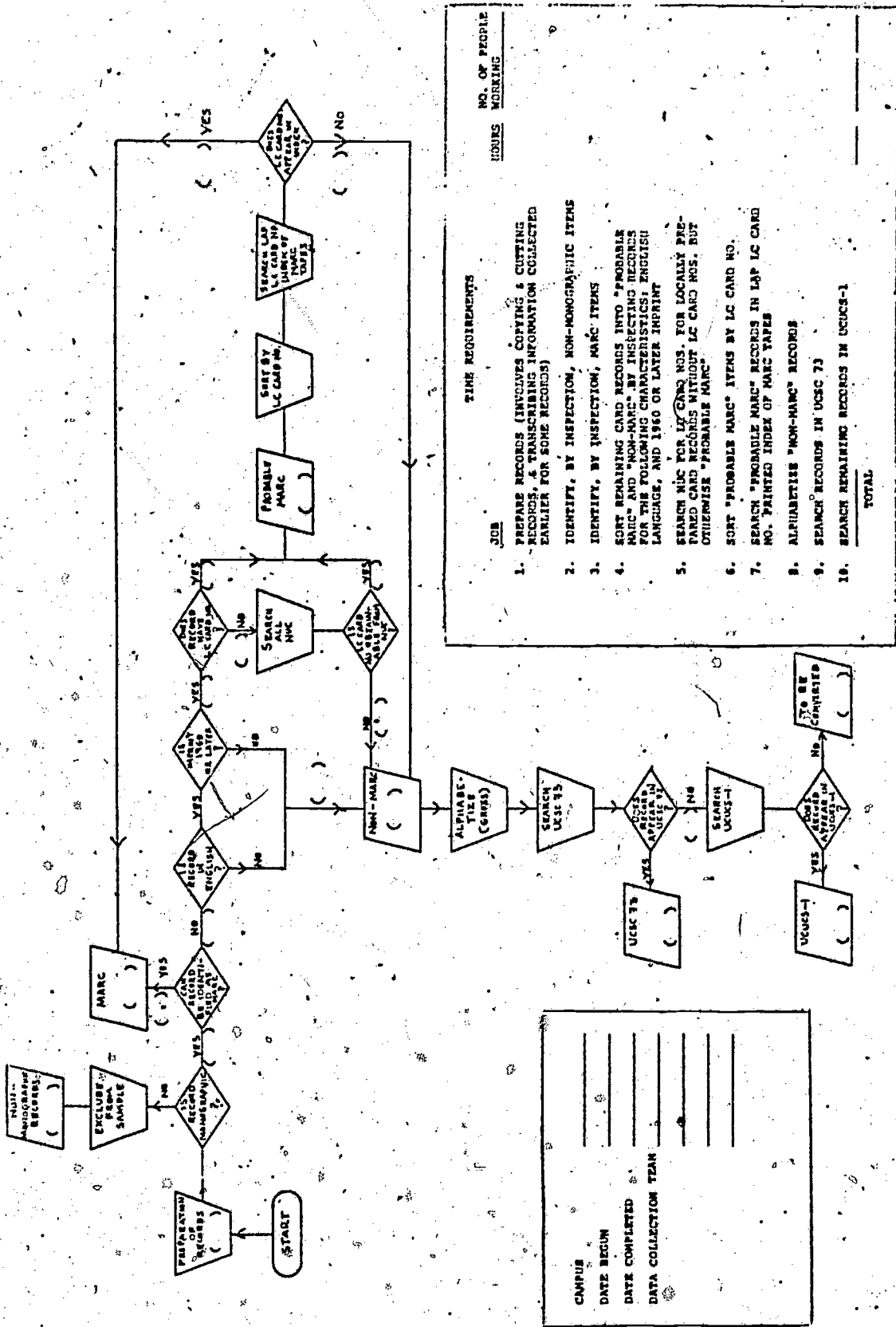


Figure 2. Flowchart of Record Search Process for UC Machine Files

<u>PUBLICATION DATE</u>	<u>NUMBER OF RECORDS WITH THIS DATE</u>	<u>PERCENT OF TOTAL</u>	<u>CUMULATIVE PERCENTAGE</u>
NO DATE	146	.03	.03
PRE-1960	16,183	2.98	3.01
1960	2,376	.44	3.45
1961	2,321	.43	3.88
1962	3,067	.56	4.44
1963	3,852	.71	5.15
1964	5,093	.94	6.09
1965	6,922	1.27	7.36
1966	19,170	3.53	10.89
1967	36,235	6.67	17.56
1968	60,052	11.05	28.61
1969	61,382	11.28	39.89
1970	63,960	11.77	51.66
1971	62,378	11.47	63.13
1972	64,561	11.87	75.00
1973	66,205	12.18	87.18
1974	49,900	9.18	96.36
1975	19,358	3.56	99.92
1976	435	.08	100.00
1979	<u>2</u>	<u>.00</u>	
	543,598	100.00	

TABLE 1

Distribution of Publication Dates in Full LC MARC Data Base
as of July 1975

records for the publication dates covered in the UCUCS-2 material. For these reasons, we make a simplifying assumption and judged any record at this point that was in a foreign language and had an imprint of 1959 or earlier to be a non-LC MARC record; all of the remaining records at this point were still "probable MARC" records.

If such a "probable MARC" record did not have an LC card number, as subsequently determined by a search in the National Union Catalog publications, it was assumed not to be in the LC MARC data base and was then transferred to the non-MARC pile.

There were several ways in which the "probable MARC" records could be resolved. The way we chose was to examine a printed LC Card Number index to the full LC MARC data base that was prepared for the UC Bibcenter file. The lookups were speeded up by first manually sorting the sample records into LC Card Number order. Records that were found in this search were then annotated with the number and transferred to the LC MARC record pile. Because the printed index was prepared by the UC Bibcenter several months prior to the time at which we were doing our searching, there was the possibility that some of the "probable MARC" records might have been added to the LC MARC data base by LC after our index was printed. In order to bring our search up to date and cover absolutely everything that was in the LC MARC data base, we then searched some of the remaining "probable MARC" records on-line against the most recent LC MARC records in the Stanford BALLOTS system. A total of nine LC MARC records were found here as the result of 215 searches. The remaining 894 "probable MARC" records were then searched against the latest copy of the MCRS microfilm index of the LC MARC data base that is distributed weekly and provides the most up-to-date index to the full LC MARC data base. A

Small number of additional MARC records were found this way and resolved all of the "probable MARC" records.

All of the non-MARC records were then given a gross alphabetic sort by the first two or three characters of the author or title entry in order to speed up the manual lookups in the printed 1973 UCSC author-title book catalog. Matching records for the UCSC data base were set aside in a separate pile.

The remaining sample records that were not found in the LC MARC or UCSC catalogs were then searched in the printed UCSCS author-title book catalog and split into piles for matches and non-matches. This completed the search process in the UC machine files.

It would be a relatively straightforward task to search the remaining non-matching (i.e., "residue") material in any of several other available machine data bases, and this in fact was done with the OCLC data base.

In all of the searches a match was recorded when there was either an exact match or a "near" match. A near match was defined as a situation in which the catalog records were identical except for a few minor changes such as edition, date, or place of publication. In one of our initial tests we found that broadening the overlap to include near matches increased the total overlap figures by about five percent.

2. Search in Ohio College Library Center (OCLC) File

Arrangements were made with OCLC to permit a terminal at ILR to be used to search the OCLC data base on-line. This was done over the Tymnet data communications network and was restricted by the Tymnet line to a transmission rate of 30 characters per second. A Texas Instruments Silent 700 thermal printer terminal was used at this 30 character per second rate to do the searching and to obtain a printed record of all of the search.

transactions. Most of the searches were done by ILR staff members with library experience, although a few searches were made by staff members with less experience. All of the searching was done during the months of January to July 1975.

It was known in advance that OCLC searching by LC Card Number (LCCN) was considerably faster than searching by title or author-title search keys. For that reason, LCCN searching was done first for any record that had such a number on the card. Cards that did not get a hit with the LCCN search or did not have an LCCN were given a title or author-title search. Because the title search was generally faster than an author-title search and yielded fewer erroneous citations in response to the search code, it was the preferred form of search for the cards that did not get a hit on LCCN searching. Cards that did not get a hit on a title search were also given an author-title search.

The final results of the search effort were then annotated on each of the sample source records.

IV. FINDINGS

A. TOTAL SOURCE RECORDS TO BE CONVERTED

UCUCS-2 cards were received in 1973 from all campuses except Santa Cruz. As noted earlier, all of the data and the conversion problem relate to the eight non-UC Santa Cruz campuses. No cards were requested or received from the more than 100 UC libraries that are not affiliated with the university library systems of each campus.

Under the direction of Tom Hargrove of ILR, all of the cards received by ILR were given some pre-processing which included re-packaging into standard size boxes, serial numbering of each card, microfilming of each card, the separation and separate boxing of records that contained non-Roman characters, and the exclusion of some records (e.g., notes and non-monograph catalog records) that were not meant for conversion. As a result of this pre-processing effort, some fairly accurate data was available from Tom Hargrove regarding the gross counts of the source records.

Because of the special and difficult problems associated with the computer representation and processing of non-Roman alphabetic information, most of the records submitted for UCUCS-1 that contained non-Roman-alphabet information were set aside for special processing and not included in UCUCS-1. These approximately 63,000 records have still not received any processing. It is assumed that the same practice will be followed for UCUCS-2, namely that the non-Roman material will be separated out and not converted with the other material. Thus only the Roman-alphabet material is of interest to the conversion problem. It should be clear, however, that some consideration eventually needs to be given to the approximately 193,000 non-Roman records (63,784 for UCUCS-1, 130,099 for UCUCS-2) that have been

submitted to date for the UC union catalog efforts.

Because the non-Roman records have already been set aside for separate processing, this study concerned itself only with planning for the conversion of the Roman-alphabet records. According to the data in Table 2, this leaves us with the problem of handling almost 1.7 million catalog cards as input to the conversion effort. Our problem is described in terms of cards rather than titles because all of the programming, numbering, linkage of continuation cards, and control of card conversion is done by a locally assigned card number rather than title. We will in fact be faced with a total number of titles which is smaller than this card count.

NUMBER OF CARDS RECEIVED

<u>CAMPUS</u>	<u>ROMAN</u> ¹	<u>NON-ROMAN</u> ²	<u>SPECIAL MATERIAL</u> ³	<u>TOTAL</u>
Santa Cruz	88,000 (tape records)	-	-	88,000 (tape records)
Los Angeles	341,472	33,734	13,790	388,996
Santa Barbara	283,632	20,742	16,284	320,658
Berkeley	266,696	53,373	325	320,394
Davis	228,762	8,407	25	237,194
San Diego	209,921	3,948	24,900	238,769
Riverside	198,987	6,361	183	205,531
Irvine	141,384	3,036	2,475	146,895
San Francisco	26,968	498	525	27,991
TOTAL CARDS:	1,697,822	130,099	58,507	1,886,428
TOTAL FILE:	1,785,822	130,099	58,507	1,974,428

NOTES

1. ROMAN--cards in all Roman-alphabet script, including diacritics.
2. NON-ROMAN--cards containing some NON-ROMAN script, even if text cataloged is in Roman script, either as transliteration or translation into a Roman-script language.
3. SPECIAL--material requiring special handling due to non-regular card stock, special notation, illegibility, special cataloging--brief, photo-listing.

TABLE 2

Materials Received for UCUCS-2

B. NATURE OF THE SOURCE RECORDS TO BE CONVERTED

The UCUCS-2 records⁸ represent most of the monographs cataloged by the UC General Libraries during the period 1968 through 1972. As such, this sample represents much of the type of material that was going into the UC collections during this recent 5-year period. For that reason it is of interest to determine the general characteristics of this material. All of the data in this section is based on a study of our entire sample of source records.

1. Language

An analysis of the language of publication of our sample monograph records shows that approximately 34% of this material was in a foreign language. Berkeley had the highest percentage (44%) of foreign language monographs. Detailed data is given in Table 3.

2. Imprint Date

The distribution of imprint dates of monograph material in the sample is given in Table 4 and illustrated in Figure 3. Almost half of the material was published during the 5-year period in which the UCUCS-2 material was collected, and about two-thirds of the UCUCS-2 material was published during the total UCUCS time period of 1963-72. About one quarter of the UCUCS-2 records were for material more than 20 years old.

3. Availability of Unique Identification Number

Suggestions have been made of the possibility of preparing a computer-based numeric register or book catalog for the UC library resources in a manner similar to that done recently for the Louisiana libraries.* Such

* McGrath, William E. and Donald Simon, "Regional Numerical Union Catalog on Computer Output Microfiche," Journal of Library Automation 5:4 (December 1972) 217-229. See also McGrath, William E., "LNR: Numerical Register of Books in Louisiana," LLA Bulletin (Louisiana Library Association) 34:3 (Fall 1971) 79-86.

<u>CAMPUS</u>	<u>SAMPLE SIZE</u>	<u>ENGLISH RECORDS</u>		<u>NON-ENGLISH RECORDS</u>	
		<u>Number</u>	<u>Percent</u>	<u>Number</u>	<u>Percent</u>
Berkeley	1,134	640	56.44	494	43.56
Santa Barbara	1,453	889	61.18	564	38.82
Riverside	1,235	797	64.53	438	35.47
Los Angeles	1,411	952	67.47	459	32.53
Davis	932	656	70.39	276	29.61
Irvine	666	487	73.12	179	26.88
San Diego	1,103	830	75.25	273	24.75
San Francisco	95	78	82.11	17	17.89
TOTAL	8,029	5,329	66.37	2,700	33.63

TABLE 3

Distribution of English and Non-English UCUCS-2 Monograph Records

B ———
LA ———
D
I ———
SB ———
SD
SF ———
R - - - -

15

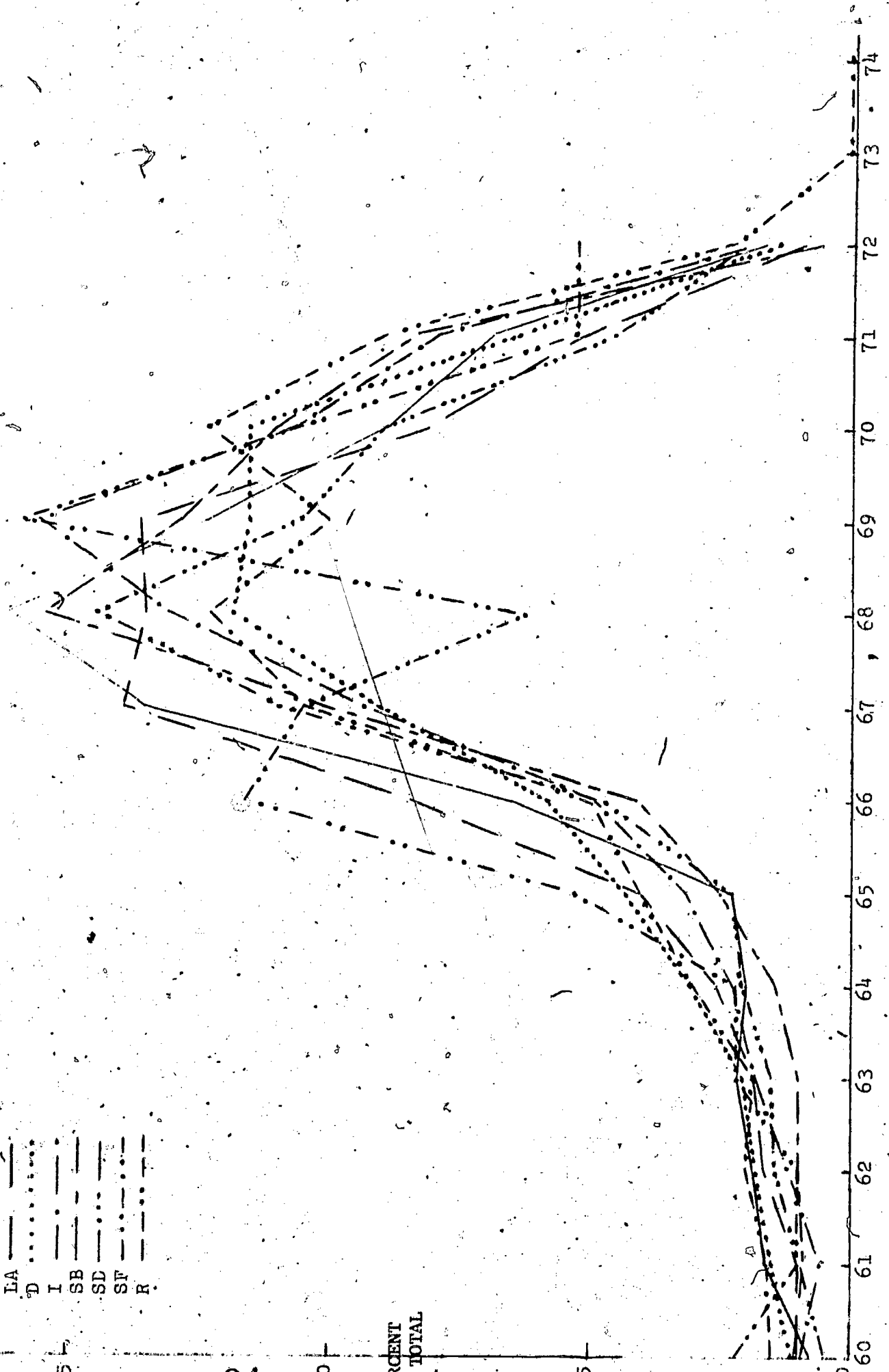
24

10

PERCENT
OF TOTAL

5

0



PUBLICATION DATE

Figure 3a. Age Distribution of UCUCS-2 Monograph Records

100
90
80
70
60
50
40
30
20
10
0

- B —————
- LA —————
- D ······
- I —————
- SB —————
- SD —·····
- SF —·····
- R —·····

PERCENT
OF TOTAL

60 62 64 66 68 70 72 74

PUBLICATION DATE

Figure 3b: Cumulative Age Distribution of UCUCS-2' Monograph Records

25



Year/Date	Berkeley		Davis		Irvine		Santa Barbara		Los Angeles		San Diego		San Francisco		Miverside		TOTAL	
	No.	CZ	No.	CZ	No.	CZ	No.	CZ	No.	CZ	No.	CZ	No.	CZ	No.	CZ	No.	CZ
1974	0	-	0	-	0	-	0	-	0	-	0	-	0	-	1	0.1	1	0.0
1973	18	1.6	13	1.4	12	1.8	10	0.7	12	0.9	21	2.1	5	5.3	20	2.3	1	0.0
1972	77	6.8	64	7.8	52	7.8	40	5.2	74	5.2	48	6.7	5	5.3	106	8.6	122	1.5
1971	100	8.8	107	11.5	107	10.7	160	11.0	114	8.1	142	13.5	10	10.5	182	12.3	585	6.8
1969	140	12.3	29.5	19.3	103	15.5	186	12.8	193	13.7	279	26.0	15	15.8	335	23.9	977	12.2
1968	183	16.1	45.6	30.8	135	20.3	224	15.4	196	13.9	313	29.4	6	6.3	432	31.2	1,114	11.9
1967	157	13.5	59.1	40.1	147	21.7	187	13.1	196	13.9	313	29.4	10	10.5	432	31.2	1,114	11.9
1966	74	6.5	59.1	40.1	59	4.9	147	10.1	186	13.9	313	29.4	11	11.6	432	31.2	1,114	11.9
1965	26	2.3	67.9	40.1	33	4.9	82	4.1	119	8.4	317	32	2	2.2	70.6	4.9	60.5	5.8
1964	23	2.0	69.9	40.1	37	3.2	58.6	3.8	37	4.0	67.7	23	2	2.2	72.8	3.1	67.6	3.2
1963	25	2.2	72.1	40.1	15	2.3	65.3	3.7	26	1.8	70.3	23	2	2.2	75.0	3.1	160	1.7
1962	20	1.8	71.9	40.1	14	1.2	71.4	1.0	26	1.8	72.1	17	1	1.0	76.0	1.8	131	1.5
1961	18	1.6	75.5	40.1	14	1.2	71.4	1.0	26	1.8	72.1	17	1	1.0	76.0	1.8	131	1.5
1960	8	0.7	76.2	40.1	6	0.6	72.0	0.8	14	1.0	74.8	10	0	0.0	77.0	1.0	93	1.1
1959	9	0.8	72.0	40.1	6	0.6	72.0	0.8	14	1.0	74.8	10	0	0.0	77.0	1.0	93	1.1
1958	9	0.8	72.0	40.1	6	0.6	72.0	0.8	14	1.0	74.8	10	0	0.0	77.0	1.0	93	1.1
1957	3	0.3	77.7	40.1	6	0.9	75.9	0.6	8	0.6	77.3	12	0	0.0	81.2	0.9	48	0.6
1956	3	0.4	78.1	40.1	6	0.9	75.9	0.6	8	0.6	77.3	12	0	0.0	81.2	0.9	48	0.6
1955	4	0.4	78.5	40.1	6	0.9	75.9	0.6	8	0.6	77.3	12	0	0.0	81.2	0.9	48	0.6
1954	2	0.2	78.7	40.1	7	1.1	76.7	0.8	12	0.8	78.8	11	1	1.0	82.2	1.1	50	0.6
1953	7	0.6	79.3	40.1	5	0.7	79.8	0.6	11	0.8	81.1	11	1	1.0	84.2	1.1	42	0.5
1952	1	0.1	79.4	40.1	4	0.6	80.4	0.5	8	0.6	81.1	11	1	1.0	84.2	1.1	42	0.5
1951	3	0.3	79.9	40.1	3	0.3	81.2	0.5	7	0.5	81.2	11	1	1.0	84.2	1.1	42	0.5
1950	2	0.2	79.9	40.1	2	0.3	81.2	0.5	7	0.5	81.2	11	1	1.0	84.2	1.1	42	0.5
1940	27	2.4	82.3	46	4.9	84.2	25	3.6	34	2.2	82.8	57	0	0.0	84.2	2.2	209	3.6
1930	21	1.9	84.7	32	3.6	87.6	34	5.1	28	2.0	86.8	49	0	0.0	84.2	2.2	209	3.6
1920	17	1.5	85.7	39	4.2	91.8	21	3.2	28	2.0	86.8	49	0	0.0	84.2	2.2	209	3.6
1910	17	1.5	85.7	39	4.2	91.8	21	3.2	28	2.0	86.8	49	0	0.0	84.2	2.2	209	3.6
1900	11	1.0	87.3	12	1.3	94.6	8	1.2	20	0.8	87.1	20	0	0.0	84.2	2.2	209	3.6
1890	32	2.8	90.1	15	1.6	96.2	9	1.4	28	1.5	89.3	21	2	2.2	84.2	2.2	209	3.6
1880	15	1.3	91.4	12	1.3	97.5	5	0.7	28	1.5	89.3	21	2	2.2	84.2	2.2	209	3.6
1860	12	1.0	92.4	5	0.5	98.0	2	0.3	13	0.9	90.6	13	1	1.0	84.2	2.2	209	3.6
1800	21	1.9	94.3	8	0.9	98.9	2	0.3	10	0.7	91.9	10	0	0.0	84.2	2.2	209	3.6
1750	16	1.5	94.8	6	0.7	99.1	1	0.1	7	0.5	93.4	7	0	0.0	84.2	2.2	209	3.6
1700	6	0.6	95.4	2	0.2	99.3	0	0.0	3	0.3	94.7	2	0	0.0	84.2	2.2	209	3.6
0-1699	34	3.0	98.4	4	0.4	99.7	1	0.1	47	3.3	98.0	3	0	0.0	84.2	2.2	209	3.6
no date or uncertain	18	1.6	100.0	3	0.3	100.0	14	1.0	10	0.7	100.0	15	0	0.0	100.0	10	0.9	100.0
TOTAL	1,134		932		666		1,433		1,411		1,103		95		1,235		72	0.9

* Cumulative percentage

TABLE 4
Age Distribution of U.S. Census-2 Monograph Records

a finding tool would consist simply of an LC Card Number or ISBN, or ISSN, along with a location code for the holding library, and would be relatively inexpensive to produce. Obviously, it could only be used as a limited finding tool and would perhaps be of some interest to library selection staffs in the context of cooperative collection development and utilization. If such a numeric register were to be considered, some information would be necessary to determine what fraction of the holdings had some kind of unique identification number that could be used as the access point in this type of catalog.

Some of the existing UC catalog records contain such a national or international number on the local catalog record. Some of the records are associated with numbers that can only be found by searching some other catalog such as the National Union Catalog (NUC); this often happens, for example, when a catalog card is prepared locally in advance of cataloging by LC, and not replaced or augmented subsequently by the LC information.

The data in Table 5 indicates that about 81% of the UCUCS-2 material had some sort of unique number associated with it, but a large fraction of the numbers were not on the local cards and would have to be found by a relatively time-consuming lookup process. About 63% of the total UCUCS-2 records had some type of unique material identification number on the local catalog record. All of the records that had an ISBN also had an LC Card Number. No ISSN entries were found even though it was theoretically possible to have an ISSN and LCCN on the same work.

Using the best available information regarding the total number of Roman-alphabet UCUCS-2 records, we see from the data in Table 6 that if a numeric register were to be constructed, it could represent about 1.06 million of these cards if the information was taken directly from the

RECORDS WITH NUMBER ON CARD

	San Diego		Irvine		Santa Barbara		Davis		Berkeley		Riverside		San Francisco		Los Angeles		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
LC Card Number	743	67.4	467	70.1	1,065	73.3	631	67.7	679	59.9	665	53.9	61	64.2	721	51.1	5,032	62.7
ISBN	111	10.1	96	14.1	197	13.6	105	11.3	78	6.9	143	11.6	1	1.1	73	5.2	802	10.0
ISSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Any of These	743	67.4	467	70.1	1,065	73.3	631	67.7	679	59.9	665	53.9	61	64.2	721	51.1	5,032	62.7

RECORDS WITH NUMBER FOUND AFTER LOOKUP

LC Card Number	219	19.9	113	17.0	199	13.7	165	17.7	195	17.2	275	22.3	10	10.5	324	23.0	1,500	18.7
ISBN	0	0	0	0	3	2	0	0	0	0	19	1.5	0	0	0	0	22	3
ISSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Any of These	219	19.9	113	17.0	199	13.7	165	17.7	195	17.2	275	22.3	10	10.5	324	23.0	1,522	18.7

TOTAL RECORDS WITH A NUMBER

LC Card Number	962	87.2	580	87.1	1,264	87.0	796	85.4	874	77.1	940	76.1	71	74.7	1,045	74.1	6,532	81.4
ISBN	111	10.1	94	14.1	200	13.8	105	11.3	78	6.9	162	13.1	1	1.0	73	5.2	824	10.3
ISSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Any of These	962	87.2	580	87.1	1,264	87.0	796	85.4	874	77.1	940	76.1	71	74.7	1,045	74.1	6,532	81.4

RECORDS WITH NO NUMBER

LC Card Number	141	12.8	86	12.9	189	13.0	136	14.6	260	22.9	295	23.9	24	25.3	366	25.9	1,497	18.6
ISBN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Any of These	141	12.8	86	12.9	189	13.0	136	14.6	260	22.9	295	23.9	24	25.3	366	25.9	1,497	18.6

TOTAL SAMPLE

LC Card Number	1,103	100.0	666	100.0	1,453	100.0	932	100.0	1,134	100.0	1,235	100.0	95	100.0	1,411	100.0	8,029	100.0
----------------	-------	-------	-----	-------	-------	-------	-----	-------	-------	-------	-------	-------	----	-------	-------	-------	-------	-------

TABLE 5
Proportion of UCICS-2 Records with LC Card, ISSN, or ISSN Numbers

NUMBER OF ROMAN-ALPHABET CARD RECORDS
RECEIVED FOR UCUCS-2

<u>CAMPUS</u>	<u>Total Records</u>	<u>Total With Number On Card</u>	<u>Total With Number Found After Lookup</u>	<u>Total With Number</u>
Los Angeles	341,472	174,492	78,402	252,894
Santa Barbara	283,632	207,902	38,858	246,760
Berkeley	266,696	159,698	45,872	205,570
Davis	228,762	154,872	40,491	195,363
San Diego	209,921	141,403	41,669	183,072
Riverside	198,987	107,154	44,314	151,468
Irvine	141,384	99,138	23,993	123,131
San Francisco	<u>26,968</u>	<u>17,316</u>	<u>2,840</u>	<u>20,156</u>
TOTAL	1,697,822	1,061,975	316,439	1,378,414

TABLE 6

Number of Records with a Unique National or International Identification Number

available cards, and a total of about 1.38 million of these records if additional lookups were to be made to search for missing numbers.

C. AVAILABILITY OF SOURCE RECORDS IN EXISTING MACHINE FILES

1. Availability in the UC Files (LC MARC, UCSC, UCUCS-1)

Because it is generally so much more expensive to create a new machine record than to copy it from some of the existing data bases of catalog records, one major file conversion policy would be to make as much use as possible of available machine data bases. Project and budget planning for UCUCS-2 record conversions will need a good estimate of the number of records that might be copied, as well as the number of records that might have to go through an original conversion process.

Using the search procedures and sequence described in the earlier section on our method of approach, the records for each campus were examined to find the extent to which they overlapped with three existing UC machine data bases: LC MARC, UCSC, and UCUCS-1. The results of this analysis of 8,337 sample records are summarized in Table 7.

As shown by the data in Tables 7 and 8, about 27% of the UCUCS-2 material can presently be taken from the LC MARC data base. Over 15% of the source cards will be immediately identified as LC MARC records by inspection of the card, and another 11% of the source cards could only be identified as LC MARC records by a bibliographic author-title search in the National Union Catalog (1956-68 cumulation, and 1968-72 cumulation).

A total of over 48% of the UCUCS-2 material can be taken from at least one of these three UC files. Some of the remainder can be taken from other files such as OCLC or SDC/Information Dynamics LIBCON.

For some of the campus samples, the records were searched against each of the three UC data bases to find the extent to which each of these data bases contributed unique records. These searches indicated that each of the three data bases contribute some records that were unique to that data

TOTAL UCUCS-2 CARD RECORDS**

SAMPLE RECORDS

CAMPUS	Total Sample Size	CARD RECORDS AVAILABLE IN UC MACHINE FILES*			Additional Records Available In OCLC	Estimated To Be Excluded From Processing	Estimated To Be Available In The 3 UC Machine Files	Additional Records Estimated To Be Available In OCLC File	Estimated To Require Original Conversion If Only UC Files Used	Estimated To Require Original Conversion If UC And OCLC Files Used		
		LC MARC	UCSC	UCUCS-1								
LA	1,617	350 (21.65)	88 (5.44)	140 (8.56)	578 (35.74)	206 (12.74)***	341,472	43,504 (12.74)	122,042 (35.74)	41,606	175,926	134,320
B	1,238	302 (24.39)	53 (4.28)	74 (5.98)	429 (34.65)	104 (8.40)	266,696	22,402 (8.40)	92,410 (34.65)	29,300	151,884	122,584
SB	1,587	449 (28.29)	134 (8.44)	194 (12.19)	777 (48.96)	134 (8.44)	283,632	23,939 (8.44)	138,866 (48.96)	39,858	120,827	80,969
D	1,015	307 (30.25)	105 (10.34)	116 (11.43)	528 (52.02)	83 (8.18)	228,762	18,713 (8.18)	119,002 (52.02)	35,833	91,047	55,214
SD	1,156	347 (30.02)	149 (12.89)	204 (17.65)	700 (60.55)	53 (4.58)	209,921	9,614 (4.58)	127,107 (60.55)	26,337	73,200	46,863
R	1,423	361 (25.37)	195 (13.70)	191 (13.42)	747 (52.49)	188 (13.21)	198,987	26,286 (13.21)	104,448 (52.49)	26,014	68,253	42,239
I	704	267 (37.93)	86 (12.22)	90 (12.78)	443 (62.93)	38 (5.40)	141,384	7,635 (5.40)	88,973 (62.93)	15,662	44,776	29,114
SF	97	26 (26.80)	4 (4.12)	23 (23.71)	53 (54.64)	2 (2.06)	26,968	556 (2.06)	14,735 (54.64)	2,779	11,677	8,898
TOTAL	8,837	2,409 (27.26)	814 (9.21)	1,032 (11.68)	4,255 (48.15)	808 (9.14)	1,697,822	152,648	807,583	217,389	737,590	580,201

* Including both exact and near matches.
 ** Excluding non-Roman alphabet cards and special material.
 *** Numbers shown in parentheses are percentages.

TABLE 7

Summary of Machine Record Availability from External Data Bases for UCUCS-2 Records

CAMPUS	TOTAL SAMPLE SIZE	ITEMS IDENTIFIABLE BY INSPECTION AS MARC RECORDS		ADDITIONAL ITEMS FOUND BY BIBLIOGRAPHIC SEARCHING TO BE MARC RECORDS		TOTAL MARC RECORDS	
		Number	Percent	Number	Percent	Number	Percent
I	704	166	23.58	101	14.35	267	37.93
SD	1,156	150	12.98	197	17.04	347	30.02
D	1,015	187	18.42	120	11.82	307	30.25
SB	1,587	308	19.41	141	8.88	449	28.29
SF	97	13	13.40	13	13.40	26	26.80
R	1,423	227	15.95	134	9.42	361	25.37
B	1,238	187	15.11	115	9.29	302	24.39
LA	<u>1,617</u>	<u>154</u>	9.52	<u>196</u>	12.12	<u>350</u>	21.65
TOTAL	8,837	1,392		1,017		2,409	
COMPOSITE %			15.75		11.51		27.26

TABLE 8

UCUCS-2 Monograph Records Identified as LC MARC Records

base. This data base overlap information is illustrated in Figures 4, 5, and 6.

The initial estimates given in Tables 2 and 7 for the size of the conversion problem can be modified slightly as a result of this detailed record-by-record examination of the 8,337 sample records. This detailed examination found that a small fraction of the numbered and counted UCUCS-2 material can be excluded from further processing. The UCUCS-2 Roman-alphabet source records, totalling 1,697,822 cards in Table 2 were assumed to be all monograph catalog records ready for conversion. However, the detailed review of our sample indicates, as shown by the data in Table 7, that a fraction of these records are actually for serial records or other material that was defined to be out of scope for UCUCS-1 material, and assumed to be out of scope for this study. The UCUCS-2 pre-processing operation removed much of this material, but some of it still slipped through. When this out-of-scope material, estimated to be about 9 percent of the total Roman-alphabet card records, is subtracted from the total received UCUCS-2 records, we have a smaller total number of records for conversion than our earlier estimate given in Table 2. This total still assumes that none of the card records will be manually consolidated before conversion.

With this minor adjustment, as shown in Table 7, we now see a total of at least 807,000 UCUCS-2 card records that can be copied from existing UC machine files.

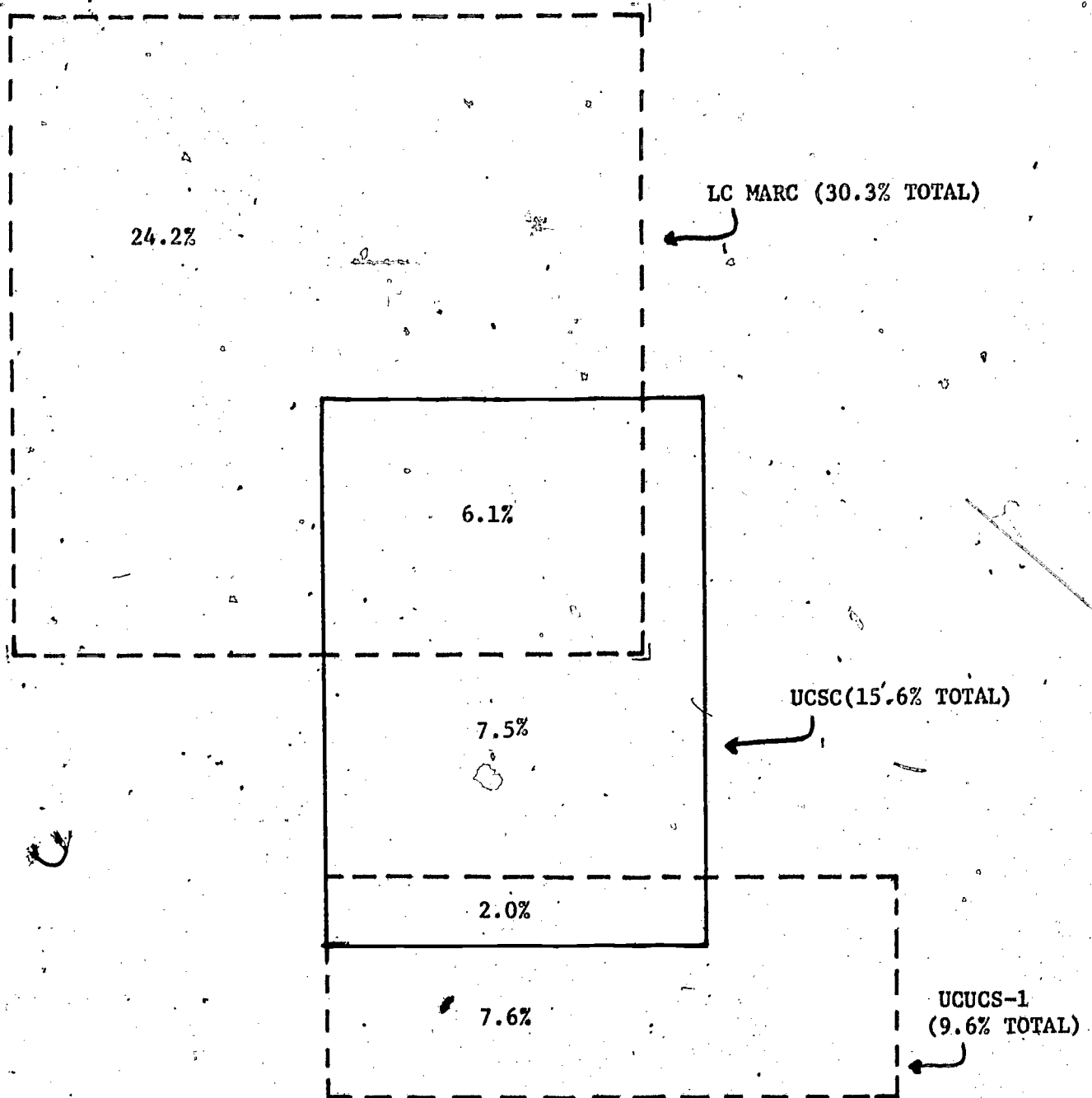


Figure 4. UC File Overlap for Riverside UCUCS-2 Records

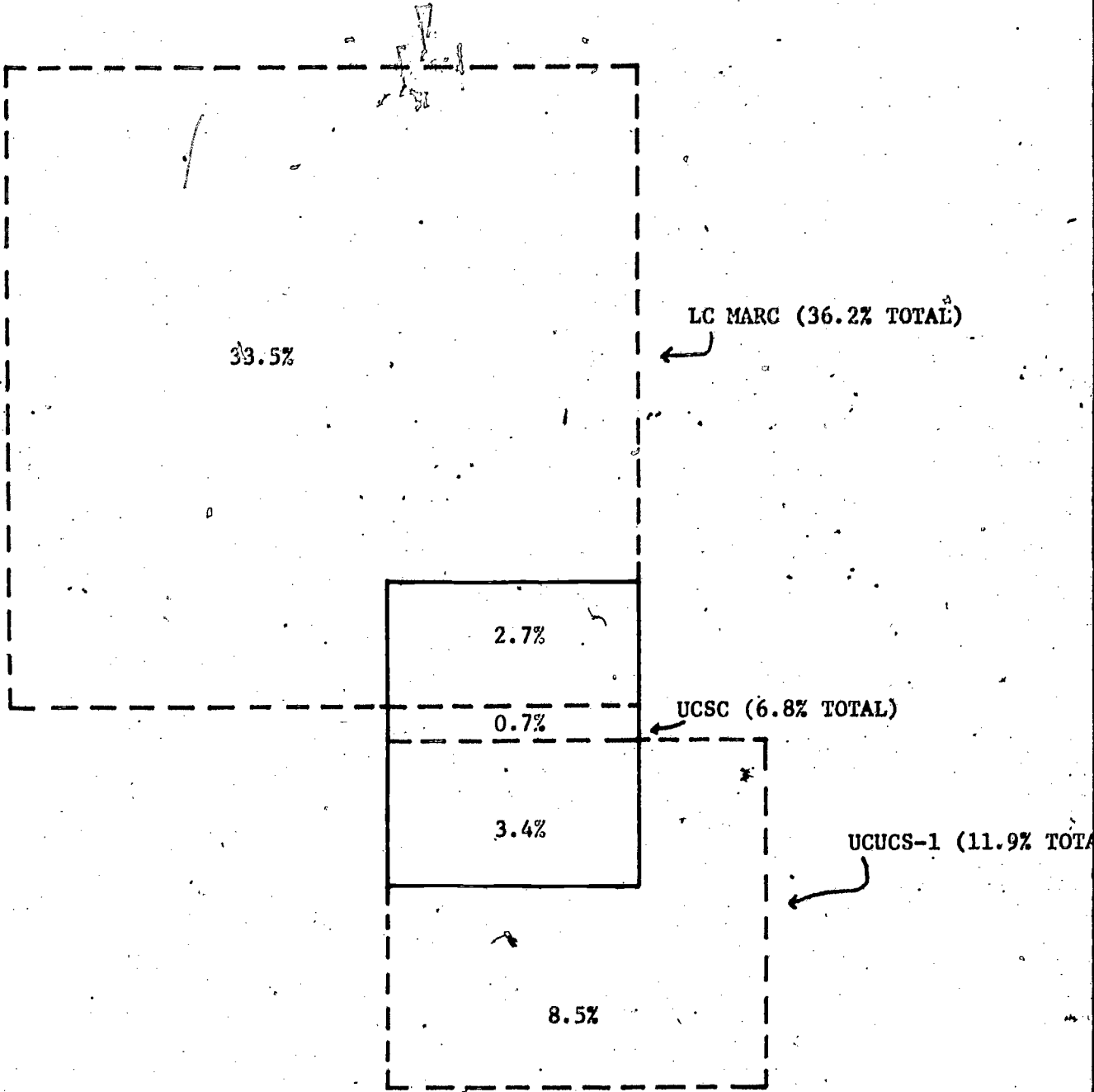


Figure 5. UC File Overlap for Davis UCUCS-2 Records

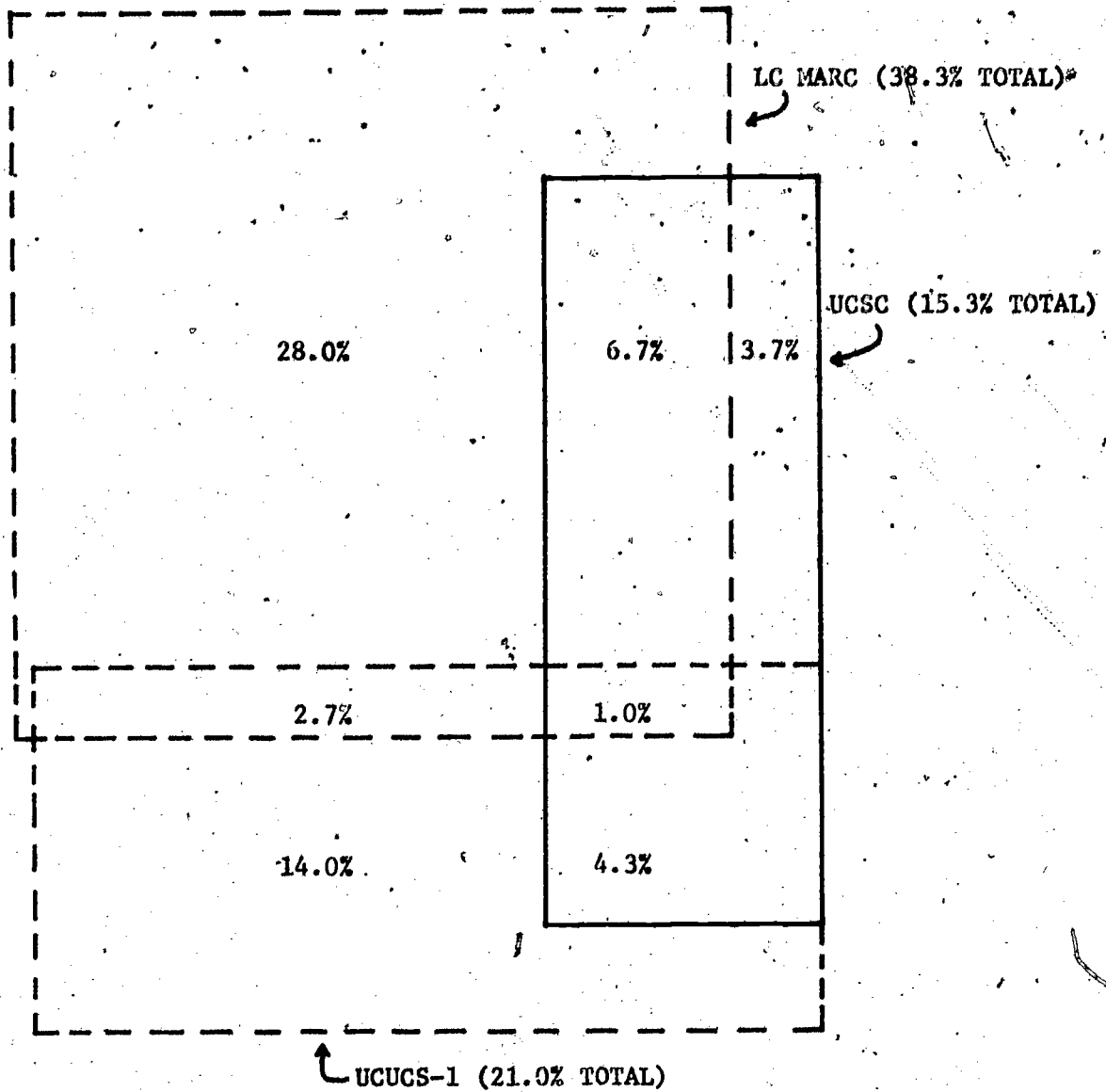


Figure 6. UC File Overlap for Irvine UCUCS-2 Records

2. Availability in the OCLC File

The Ohio College Library Center (OCLC) presently has over one million catalog records available for on-line computer searching and copying. Approximately 400,000 of these records duplicate the LC MARC file that UC already maintains.

All of the UCUCS-2 records that were not found in any of the three UC files were searched against OCLC.

In summary, as shown by the data in Table 9, a total of 30% of the residue records (i.e., those records not found in any of the three UC data bases) were found in the OCLC data base. Extrapolating this to the total UCUCS-2 card file, we find that after first searching against the three available UC files, an additional 217,000 records could be found in the OCLC file.

RESIDUE¹ SAMPLE

TOTAL RESIDUE

CAMPUS	Total Number in Residue Sample	Number of Residue Sample in OCLC	Percent of Residue Sample in OCLC	Estimated Total Number of Residue Records	Estimated Number of Residue Records in OCLC
Los Angeles	883	197	23.6	175,926	41,606
Santa Barbara	676	223	33.0	120,827	39,858
Davis	404	159	39.4	91,047	35,833
Berkeley	705	136	19.3	151,884	29,300
San Diego	403	145	36.0	73,200	26,337
Riverside	488	186	38.1	68,253	26,014
Irvine	223	78	35.0	44,776	15,662
San Francisco	42	10	23.8	11,677	2,779
TOTAL	3,774	1,134	30.0	737,590	217,389

1. The residue is defined as those records that could not be found in any of the 3 UC files (LC MARC, UCUCS-1, UCSC).

TABLE 9

Number of Residue Records Included in OCLC Data Base

3. Availability in Other Files

Several other data bases exist that should be examined to determine the extent to which they could contribute additional records to assist the UCUCS-2 record conversion effort. Files that already exist and merit additional examination are:

- . SDC/Information Dynamics LIBCON (1.2 million index entries, but a small number of non-LC MARC records)
- . BALLOTS (for the several thousand locally-generated records)
- . Blackwell North America, Inc.
- . Auto-Graphics
- . New York Public Library
- . General Research Corp.
- . Information Design

Several other files that may be prepared in the near future should also be examined, including:

- . California State Library (planned conversion of union catalog)
- . CSUC shelf list conversion.

RECENT ILR PUBLICATIONS

Publications of papers and reports of interest to scholars and practitioners in the field of library and information science is an important function of the Institute of Library Research. In addition to this study, the following have been published recently by ILR:

- ILR-73-001 Todd, Judy, Summary Report of Student Studies of the Subject Headings Used in the University of California, Berkeley, Subject Catalog (July 1973) 8 pp. (ERIC No. ED-082 775)
- ILR-73-002 Bourne, Charles P., and Jo Robinson, SDI Citation Checking as a Measure of the Performance of Library Document Delivery Systems (July 1973) 10 pp. (ERIC No. ED-082 774)
- ILR-73-003 Weeks, Kenneth, Determination of Pre-Acquisition Predictors of Book Use: Final Report (July 1973) 20 pp. (ERIC No. ED-082 776)
- ILR-73-004 Weeks, Kenneth, Proposal for a University of California/California State University and Colleges Inter-Segmental Machine Readable Library Patron Card (August 1973) 21 pp. (ERIC No. ED-082 777)
- ILR-73-005 LeDonne, Marjorie, "Summary of Court Decisions Relating to the Provision of Library Services in Correctional Institutions," Association of Hospital and Institution/Libraries Quarterly (Winter/Spring 1973) 9 pp.
- ILR-73-006 Theilin, John, and Bonnie F. Shaw (eds.), Institute of Library Research Annual Report: July 1972 to June 1973 (September 1973) 30 pp. (ERIC No. ED-086 169)
- ILR-73-007 Dekleva, Borut, Uniform Slavic Transliteration Alphabet (USTA) (October 1973) 82 pp. (ERIC No. ED-086 164)
- ILR-73-008 LeDonne, Marjorie, Findings and Recommendations. Volume I, Survey of Library and Information Problems in Correctional Institutions (January 1974) 88 pp. (ERIC No. ED-095 842)
- ILR-73-009 LeDonne, Marjorie, Access to Legal Reference Materials in Correctional Institutions. Volume II, Survey of Library and Information Problems in Correctional Institutions (January 1974) 89 pp. (ERIC No. ED-095 843)
- ILR-73-010 LeDonne, Marjorie, David Christiano, and Jane Scantlebury, Current Practices in Correctional Library Services: State Profiles. Volume III, Survey of Library and Information Problems in Correctional Institutions (January 1974) 68 pp. (ERIC No. ED-095 844)
- ILR-73-011 LeDonne, Marjorie, David Christiano, and Joan Stout, Bibliography. Volume IV, Survey of Library and Information Problems in Correctional Institutions (January 1974) 28 pp. (ERIC No. ED-095 845)
- ILR-73-012 Gregor, Dorothy, Feasibility of Cooperative Collecting of Exotic Foreign Language Serial Titles among Health Sciences Libraries in California (February 1974) 44 pp. (ERIC No. ED-104 407)
- ILR-74-001 Nozik, Barbara, The Use Status of Books Requested from the University of California, Berkeley, Inter-Library Loan (March 1974) 11 pp. (ERIC No. ED-104 411)
- ILR-74-002 Bourne, Charles P., Institute of Library Research Annual Report: July 1973 to June 1974 (1974) 25 pp. (ERIC No. ED-097 884)
- ILR-74-003 Humphrey, Allan J., Survey of Selected Installations Actively Searching the ERIC Magnetic Tape Data Base in Batch Mode Volume I (June 1973) 86 pp. (ERIC No. ED-096 982) Volume II (June 1973) 268 pp. (ERIC No. ED-096 983)
- ILR-74-004 Cooper, William S., Donald T. Thompson, and Kenneth R. Weeks, The Duplication of Monograph Holdings in the University of California Library System (October 1974) 32 pp. (ERIC No. ED-097 883)
- ILR-74-005 Bourne, Charles P., Jo Robinson, and Judy Todd, Analysis of ERIC On-Line File Searching Procedures and Guidelines for Searching (November 1974) 140 pp. + appendices. (ERIC No. ED-101 757)
- ILR-75-001 Martell, Charles R., Jr., Interlibrary Loan Turnaround Time: A Study of Performance Characteristics of the University of California, Berkeley, Inter-Library Loan Lending Operation (January 1975) 34 pp. (ERIC No. ED-104 413)
- ILR-75-002 Bourne, Charles P., and Dorothy Gregor, Methodology and Background Information to Assist the Planning of Serials Cancellations and Cooperative Serials in the Health Sciences (January 1975) 60 pp. (ERIC No. ED-104 409)
- ILR-75-003 Bourne, Charles P., Dale Read, and Margaret Buss, Bibliographic Access to the University of California Library Resources at Berkeley and Los Angeles (June 1975) 188 pp.
- ILR-75-004 Martell, Charles R., Jr., Document Availability and Use Patterns at the University of California, Berkeley, Library: A Comparison with California State University, Sacramento (Sponsored by the General Library, University of California, Berkeley) (July 1975) 34 pp.
- ILR-75-005 Bourne, Charles P., Institute of Library Research Annual Report: July 1974 to June 1975 [1975] 60 pp.