

**THE RECON PILOT PROJECT: A PROGRESS REPORT
NOVEMBER 1969 - APRIL 1970**

Henriette D. AVRAM, Kay D. GUILLES, Lenore S. MARUYAMA: MARC Development Office, Library of Congress, Washington, D. C.

A synthesis of the second progress report submitted by the Library of Congress to the Council on Library Resources under a grant for the RECON Pilot Project. An overview of the progress made from November 1969 to April 1970 in the following areas: production, Official Catalog comparison, format recognition, research titles, microfilming, investigation of input devices. In addition, the status of the tasks assigned to the RECON Working Task Force are briefly described.

INTRODUCTION

An article was published in the June 1970 issue of the *Journal of Library Automation* (1) describing the scope of the RECON Pilot Project (hereafter referred to as RECON) and summarizing the first progress report submitted by the Library of Congress (LC) to the Council on Library Resources (CLR).

RECON is supported by the Council, the U.S. Office of Education, and the Library of Congress. In order that all aspects of the project might be brought together as a meaningful whole, the various segments, regardless of the source of support, were covered in the second progress report and have been included in this article. In some instances, it has been necessary to introduce a section by repeating some aspects already reported in the June 1970 article in order to add clarity to the content of that section.

PROGRESS—NOVEMBER 1969 TO APRIL 1970

RECON Production

The production operations of the RECON Pilot Project are being handled by the RECON Production Unit in the MARC Editorial Office of the LC Processing Department. Printed cards with 1968, 1969, and 7-series card numbers have been provided from the Card Division stock for RECON input, and approximately 99,550 cards in the 1969 and 7-series have been received. Using prescribed selection criteria the RECON editors have sorted these cards and obtained approximately 27,150 eligible for RECON input. Approximately 150,000 cards in the 1968 series have also been received. The RECON editors have sorted 60,000 of these cards and obtained approximately 24,000 records eligible for RECON input. A large number of cards in these three series is already out of print, and replacement cards are being sent by the Card Division as soon as reprints are made.

Each card eligible for RECON input from the above-mentioned selection process is also checked against a computer produced index of card numbers for records in machine readable form. Each number in the print index has a corresponding code to show on which machine readable data base the record resides. The source codes are as follows:

- M1—MARC I data base
- M2—MARC II, 1st practice tape
- M3—MARC II, 2nd practice tape
- M4—MARC II data base
- M5—MARC II residual data base

(The two practice tapes contain records converted before the implementation of the MARC Distribution Service to test the programs and input techniques.) The print index used for the final selection of the 1969 and 7-series card numbers contained only the records from M2-M5 (the MARC I data base consists of the records converted during the MARC Pilot Project which ended in June 1968). For the selection of the 1968 records, another print index had been produced which contains numbers for records on all five data bases.

If the RECON editors find a match on the print index, the appropriate source code is added to the printed card; these printed cards are then maintained in a separate file. (Later in the project, the records in the data bases identified as M1 to M3 will be updated to conform with the current MARC II format and added to the RECON data base.) The remaining cards for RECON are reproduced on input worksheets and edited. To date, approximately 9,750 records in the 1969 and 7-series have been edited for RECON.

RECON records in the 1969 and 7-series are being input by a service bureau. The contractor uses IBM Selectric typewriters equipped with an OCR typing mechanism, and the hard-copy sheets are run through an

optical scanner. The output from the scanner is a magnetic tape which is processed by the contractor's programs to produce a tape in the MARC Pre-Edit format. This tape is then sent to LC and processed by the MARC System programs to produce a full MARC record.

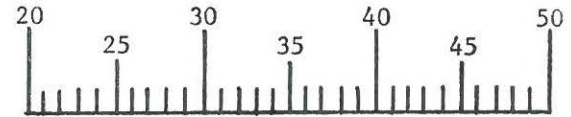
Since the input for the retrospective conversion effort will be printed cards (or copies of printed cards from the Card Division record set), it will be necessary to compare these with their counterparts in the LC Official Catalog. The printed card for each main entry in the Official Catalog will show if any changes have been made which did not warrant reprinting these cards to incorporate these changes. Items on a printed card that could be noted in this fashion include changed subject headings, added entries, and call numbers. Since these will be important access points in a machine readable catalog record, it was felt that such revisions should be reflected in the RECON records.

The RECON Report (2) contains a lengthy discussion of the various factors involved in the catalog comparison process, such as the percentage of change in relation to the age of the record, the difficulty in ascertaining any changes because of language, interpretation of cataloging rules, etc. To determine the most efficient and least costly method of catalog comparison, two RECON editors were assigned to conduct an experiment to test eight different methods as follows:

- 1) Print-out checked in alphabetic order—single group of 200 records.
- 2) Proofsheets (already proofed) checked in worksheet (card number) order—group of 200 records in batches of 20.
- 3) Proofsheets (not proofed) checked in worksheet (card number) order—group of 200 records in batches of 20.
- 4) Proofsheets (already proofed) checked by mental alphabetization—group of 200 records in batches of 20.
- 5) Proofsheets (not proofed) checked by mental alphabetization—group of 200 records in batches of 20.
- 6) Worksheets before editing (not input) checked by mental alphabetization—group of 200 records in batches of 20.
- 7) Worksheets before editing (not input) checked in alphabetical order—group of 200 records in batches of 20.
- 8) Worksheet before editing (not input) checked in worksheet (card number) order—group of 200 records in batches of 20.

Mental alphabetization means the searching of all the entries in a batch beginning with "A," then all the entries beginning with "B," etc., even though the batch is not in alphabetical order. Each editor used 200 records for each method, made the necessary corrections, and recorded the time required as well as the number of corrections made.

Figure 1 shows the average number of records checked in an hour using the eight different methods of catalog comparison. Tables 1 and 2 give the estimated cost per record for each of the methods. In determining



- Method One : PRINT-OUT Checked in ALPHABETICAL Order
- Method Two : PROFSHEETS (Already Proofed) Checked in WORKSHEET Order
- Method Three: PROFSHEETS (Not Proofed) Checked in WORKSHEET Order
- Method Four : PROFSHEETS (Already Proofed) Checked by MENTAL ALPHABETIZATION
- Method Five : PROFSHEETS (Not Proofed) Checked by MENTAL ALPHABETIZATION
- Method Six : WORKSHEETS Before Editing (Not Input) - MENTAL ALPHABETIZATION
- Method Seven: WORKSHEETS Before Editing (Not Input) - ALPHABETICAL Order
- Method Eight: WORKSHEETS Before Editing (Not Input) - Checked in WORKSHEET Order

Fig. 1. Catalog Comparison/AVERAGE Number of Records Per Hour.

manpower costs, the average salary level of Catalog Editors GS 5, 6, 7, was chosen. Also cost was based on the assumption that "one could not realistically expect peak production rates to be maintained through an eight-hour working day. Six hours were taken as the period of effective daily production to allow for training, rest periods, problem resolution, fatigue, and irregularities in work flow." (These guidelines were taken from pp. 86, 88, and 94 of the RECON Report.) On the basis of these tables it was decided to implement method 8) (worksheets before editing, not input, checked in card number order) because this appeared to be the most efficient in terms of overall requirements. Although method 1) was the fastest, it would require an additional computer run and a significant modification to the current MARC System.

Table 1. Catalog Comparison—Cost per Record

	<i>Average</i>
Method One	\$.087
Method Two	.132
Method Three	.132
Method Four	.220
Method Five	.194
Method Six	.110
Method Seven	.132
Method Eight	.100

Table 2. Adjusted Cost Figures for Catalog Comparison by Method

<i>Method</i>	<i>Average number of records per hour</i>	<i>Unadjusted cost from Catalog Comparison Study</i>	<i>Additional costs from RECON Report</i>			<i>Total adjusted cost</i>
			<i>Annual and sick leave</i>	<i>Supervision</i>	<i>Fringe benefits</i>	
4	20	\$.220	\$.037	\$.104	\$.027	\$.388
3 and 7	33	\$.132	\$.024	\$.063	\$.016	\$.235
8	44	\$.100	\$.017	\$.047	\$.012	\$.176
1	50	\$.087	\$.016	\$.041	\$.011	\$.155

The fact that a record has not been certified, i.e., that the Official Catalog card was not found, will not prevent its being added to the RECON master data base. The absence of any catalog certification information will allow these records to be isolated and searched in the Official Catalog at a later date. Catalog certification information will be added only to the LC internal format and will not be included in the communications format. Catalog comparison will also be done for the 5000 research titles to learn what problems are involved in checking older cataloging records and foreign language titles. This process will be monitored constantly in order to achieve the most economical and efficient procedures.

Format Recognition

The Library of Congress is developing a technique called "format recognition," which will allow the computer to process unedited catalog records to determine the proper content designators by examining the data string for certain keywords, significant punctuation, and other clues. The manual editing process in which the content designators are assigned is a detailed and tedious task, and it was felt that there would be considerable savings in cost by shifting some of the work to the machine for both the current MARC production and any retrospective conversion project. The background and tasks involving the format recognition process have been described in the *Journal of Library Automation* (3) and will not be repeated here. A report which will describe the format recognition algorithms in detail will be issued as a separate document in the near future.

During this reporting period, Task 2, consisting of the design and algorithms of the entire format recognition process, and Task 4, consisting of the detailed flowcharts at the coding level, based on the results of Task 2, were completed. Task 3, the extension of format recognition to foreign language records, is still in progress.

The overall systems design also involved a manual simulation process to test the general efficiency of the programs. The five people assigned to this test "processed" records following the format recognition algorithms. The unedited worksheets were input via the MT/ST by MARC typists according to the specifications that were created for format recognition. The simulators used the hard-copy output from the MT/ST for the experiment and added the machine labels to the hard copy.

All the records chosen for the simulation experiment were English language monographs but were generally "difficult", so that the error rate was considerably higher than that which would be encountered in a normal situation. There were 36 records with no errors; the remaining 114 records had one or more errors amounting to a total of 196. Of the 196 errors, 48, or approximately 24%, were made in the assignment of tags. In the present MARC System, errors in the assignment of tags require that the entire field be input again. When the format recognition process is implemented, it is likely that some adjustment will also have to be made to the present correction procedures.

After the analysis of the manual simulation was completed, the hard-copy with the tags and other content designators assigned by the format recognition algorithms was given to the input typists to convert the records for processing by the existing MARC programs. The print-out from the format recognition copy was then given to editors to proof read and correct to identify problems. This portion of the task is still being analyzed.

As the work on format recognition progresses, it becomes evident that the success of such a program depends heavily on standard cataloging practices in recording data in a certain order and in using standard punctuation. The format recognition programs being developed by the Library

of Congress have been designed to accept cataloging data based on the *Anglo-American Cataloging Rules*, but several modifications were necessary to accommodate the cataloging records created by the Shared Cataloging Division of the Library, which uses entries from various national bibliographies and adapts these for the LC printed cards.

A development with great implications for the format recognition process is the Standard Bibliographic Description (SBD). As a result of the International Meeting of Cataloging Experts held in Copenhagen in August 1969, a working party was appointed to prepare a draft proposal for the Standard Bibliographic Description. The ultimate objective is to formulate specifications for a bibliographic description in terms of order of data elements and punctuation. Use of the SBD by national bibliographies and cataloging agencies would aid interpretation of cataloging data by humans and format recognition programs. It would also aid in the exchange and transmission of cataloging data in machine readable form. The RECON Pilot Project Director, a member of this working party, is working very closely with other staff members at LC in the preparation of recommendations to the proposal.

RECON Research

One of the important aspects of the RECON Pilot Project is the conversion of 5000 records to machine readable form for research purposes. These titles would include records for English language monographs cataloged before 1950 and foreign language material (in roman alphabets only). The research titles would be used to test various input techniques and certain aspects of the format recognition program. The older material should also reveal problems because of earlier cataloging rules or differing printed card formats.

Approximately 1800 titles have been selected from the LC Main Reading Room reference catalog. This catalog is being converted to machine readable form concurrently with the RECON efforts. Additional titles to make up a total of 5000 are being selected from cards drawn by the Card Division for the RECON Production Unit. Since these cards were not eligible for RECON input, i.e., represented titles other than English language monographs, they constitute a readily available stock of printed cards to examine. The emphasis on selection criteria for these cards will be for foreign language monographs. Analysis of these research titles is also taking place to identify and possibly solve some of the problems before actual editing is begun. Many of these problems will have to be considered by the LC Processing Department to obtain decisions concerning cataloging policies.

Another research project is the investigation of microfilming techniques. The RECON Report recommended using the Card Division record set as the file from which to obtain records for conversion. Since the file is a working file, the Report recommended microfilming as the least disruptive

method of securing records for conversion. The investigation of micro-filming techniques and their costs was begun through discussions with staff members of the Library's Photoduplication Service.

To obtain valid cost estimates, it was necessary to provide the Photoduplication Service with a precise number of records to be microfilmed. Since the RECON Report recommended as first priority the conversion of records representing English language monographs from 1960 to date, this was the category chosen for obtaining microfilming cost estimates. In addition to being the first priority, this category is of such size as to make the task of filming, selecting, and preparing the records a manageable one that can be accomplished within a reasonable length of time. However, prior to the conclusion of the RECON Pilot Project, it is expected that the estimated costs of microfilming will be obtained for all the categories discussed in the RECON Report (4).

Use of the record set presents certain problems. It is arranged by card series (year) and by sequential number within each series. One can readily divide the file according to one-year periods from 1898 to 1968, but this card numbering system makes it more difficult to divide the file according to language or type of material. Therefore, if one is interested only in records representing monographs in English, there are two alternatives: a selection process prior to filming, allowing one to film less; or a selection process after filming, requiring one to film more. The RECON Report recommended dividing the record set into categories of conversion priority, filming the cards, and then reassembling the file (select first and film less). In obtaining the present microfilming cost estimates, the opposite was assumed: film all the records within the time period described and select the appropriate records afterwards (film more and select after).

The immediate advantage of the latter method is that the figure for the number of cards printed for the period 1960-1967 is firmer than the figure for the number of records representing English language monographs produced during the same period. While it might be feasible to assume a range of figures for the number of records representing English language monographs upon which to base estimates, it seemed more prudent at this point in time to use a firm figure to achieve reliable estimates. Therefore, the second method was assumed. Other advantages are: 1) this method does not require disrupting the arrangement of a working file; 2) records can be filmed as found with a minimum of intervention by the operator; and 3) after the conversion process, the microfilm can be retained as a security copy of the record set.

Figures on annual card production obtained from the Library's Technical Processes Research Office show that 840,387 cards were produced between 1960 and 1968. Using this figure as a basis, the Photoduplication Service was able to provide cost estimates for four different methods of micro-filming the 1960-1967 portion of the record set. It was assumed that the work could be done in the Card Division during the regular forty-hour week.

Table 3. *Techniques for Obtaining Records from Record Set*

Camera	Film	Reduction	Position	Feed	Paper Stock	Paper Size	Trim	Rate per Exposure (Microfilm)	Rate per Exposure (Print)	Total
(1) Rotary	16 mm. Neg.	20X	1A	Automatic	20 lb. sulphite	Approx. 3 x 5"	Hand	.0025	.065	.0675
(2) Planetary	16 mm. Neg.	9X	1A	Hand	20 lb. sulphite or 28 lb. ledger	Approx. 3 x 5"	Automatic	.025	.03	.055
(3) Planetary	16 mm. Neg.	16X	1B	Hand	20 lb. sulphite	8 x 10 1/2 w/3 x 5" card cent.	Automatic	.026	.07	.096
(4) Rotary	16 mm. Neg.	20X	1A	Automatic	Reader/ Printer Zinc Coated	Approx. 8 x 10"	Hand	.0025	.04 (2 cds. each)	.0425

If overtime were required, 40% would have to be added to the rates. The quotations given are good for one year only; i.e., the prices quoted can only be expected to prevail over the next twelve months. Beyond that time any quotation given is likely to be higher because of the general trend of rising costs. Aside from a few minimum limitations, delivery schedules for records filmed are flexible, that is, all the records need not be filmed at one time. They may be filmed in blocks, e.g., one year's records, to accommodate the requirements of the conversion unit.

Table 3 shows the components and costs for each of four different techniques that might be used in obtaining records from the Card Division record set.

Following are explanations of some of the components in the table and brief descriptions of techniques.

Cameras:

Planetary (Flat-Bed)—"A type of microfilm camera in which the document being photographed and the film remain in a stationary position during the exposure. The document is on a plane surface at time of filming." (5)

Rotary (Flow)—"A type of microfilm camera that photographs documents while they are being moved by some form of transport mechanism. The document transport mechanism is connected to a film transport mechanism, and the film also moves during exposure so there is no relative movement between the film and the image of the document." (6)

Film:

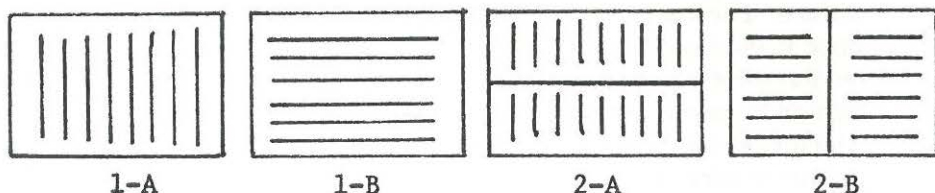
In all four techniques the film used is 16 mm. negative microfilm.

Reduction Ratio:

"The ratio of the linear measurement of a document to the linear measurement of the image of the same document expressed as 16:1, 20:1, etc." (7)

Position:

The orientation of images on a roll of film which can be controlled by turning either the document or the camera head and adjusting the reduction ratio accordingly (8).



Feed:

The method of transporting the document to be filmed past the camera head.

Paper Stock:

The type of paper used in restoring images to readable hard copy.

Trim:

Refers to method of cutting hard copy into sheets.

Rate per exposure (microfilm):

Unit price per image for microfilming.

Rate per exposure (print):

Unit price per image for restoring film to readable hard copy.

Brief descriptions of the four techniques shown in Table 3:

- 1) A rotary camera with 16 mm. negative microfilm is used to film the document at a reduction rate of 20:1 and includes an inspection for technical quality only. Hard copy is produced by using Xerox Copyflo printers. Copyflo enlargers print from roll microfilm on a continuous web of paper at a rate of twenty paper feet per minute. The size of the reconstituted document is approximately 3" x 5". Hard copy produced from a rotary camera must be trimmed manually.
- 2) This technique uses essentially the same process as 1) except that the filming is done by a planetary camera at a reduction ratio of 9:1 and includes 100% inspection for completeness and technical quality. For the planetary camera, a cutting line is provided on the film. This line enables the subsequent hard copy to be machine trimmed, with considerable savings in time and labor. A rotary camera does not have this capability, which is the single factor that accounts for most of the decrease in unit price per print exposure as compared with 1).
- 3) This technique uses a planetary camera microfilming at a reduction ratio of 16:1 and includes 100% inspection for completeness and technical quality. As in 1) and 2), hard copy is produced by using the Xerox Copyflo process. The document to be filmed, the catalog card, is placed within an outline of the RECON Worksheet. Both the Worksheet and the catalog card are filmed simultaneously; the resulting 8" x 10½" hard copy is a RECON Input Worksheet immediately available for editing without further preparation.
- 4) A rotary camera microfilming at a reduction ratio of 20:1 is used in this technique and includes an inspection for technical quality only. Hard copy is produced using a reader printer. The paper stock used in producing hard copy from reader/printers has a chemically coated surface. The unit price of \$.04 per print exposure is quite tentative, since it does not include the price of either rental or purchase of the reader/printer, or the cost of labor. In addition, this technique introduces several other factors that remain to be investigated, factors which bear on the feasibility of using this technique, as well as factors that may affect the eventual unit cost.

Investigation of Input Devices

The investigation of input devices included: 1) a continuation of the monitoring of the state of the art of input devices by contact with various manufacturers and attendance at conferences; 2) preliminary investigation of the Keymatic Magnetic Tape Unit Model 1093; 3) preliminary investigation of the CompuScan Optical Character Reader; and 4) production analysis of 1969 RECON titles using an IBM Selectric typewriter with an OCR type font sphere and a Farrington Scanner, Model 3050. Each of the above is summarized in the remainder of this section.

The devices monitored have been categorized as follows: 1) key-to-cassette (requires a converter to go from cassette to computer-compatible magnetic tape); 2) key-to-computer-compatible magnetic tape; 3) key-to-magnetic-tape system (several input devices used simultaneously to generate computer-compatible magnetic tape); 4) key-to-disk system.

The factors under consideration during the investigation of each device are: 1) type of keyboard (the arrangement and the number of characters available); 2) type of display for human readability (CRT, backlight, projection, light emitting diodes, BCD, and printed hard copy); 3) record length (number of characters considered a record on magnetic tape); 4) price.

The majority of devices available today do not satisfy the requirements for the input of bibliographic data, primarily because of the limitation in the number of available characters. Table 4 is a summarization of the devices monitored.

The following definitions amplify the content of the table. A key-to-magnetic-tape system is used to mean a number of input devices sharing centralized "electronics." The "electronics" may act as a routing and recording device from input station to magnetic tape, or the "electronics" may include a mini-computer with the facility to perform many functions, such as editing, formatting, etc. In either case, one characteristic of this system is the ability to handle a large number of input devices simultaneously. In contrast to the key-to-magnetic-tape system, the devices categorized key-to-computer-compatible-magnetic-tape include those devices that share centralized "electronics" called a pooler, which handles fewer input devices simultaneously, and records the information from these devices on one magnetic tape.

The primary attraction of the keymatic magnetic tape unit is the ability to encode 256 unique characters without the use of an escape code. In addition, the layout of the keyboard is designed according to the user's specifications. The MARC character set, consisting of 175 characters, could be assigned keys in clusters. Special characters and diacritics might reside in a particular area of the keyboard, with upper- and lower-case alphabetic characters in another area. Each cluster could have a delimiter assigned to it for ease of use by the typist. Common "words" such as MARC tags

Table 4. *Input Devices*

Manufacturer	Machine Type	Model	Keyboard Configuration	Display	Record Length in Characters	Price		Remarks
						Purchase	Monthly Rental	
Cybercom	K/C	MARK1	KP	None	80	\$7970	\$145	Converter—\$180/month
Data Action	K/C	150	KP	Projection	720	\$5900	\$155	Converter—\$575/month
IBM	K/C	50	KP	Back-light	720	\$9605	\$175	Converter—\$340/month
IBM	K/C	MTST V	T	Printed	Infinite		\$100	Converter—\$340/month
		IV	T	Printed	Infinite		\$277	
Sycor	K/C	301	T	CRT	216	\$7000	\$150	Converter—\$130/month
Tycore	K/C	8500	KP	Light-Emitting Diodes	240	\$6000	\$120	Converter—\$220/month
Viatron	K/C	21	T/KP	CRT	Infinite	\$1920	\$39	Many options affecting price
Burroughs	K/M	N-7000	KP	Projection	160	\$8400 to \$12,200	\$165 to \$277	
Honeywell	K/M	Keypape	T/KP	Back-light	80-400	\$7500 to \$33,000	\$148 to \$735	Pooler for 2 stations—\$200/month extra
Keymatic	K/M	1091	T	Back-light	Infinite	\$8750	\$166	Price is for basic 88 keys. 256 unique keys available as well as optional printer.
MAI	K/M	100-92	KP	Projection	100 or 200	\$6400	\$160	Pooler for up to 8 stations—\$40/month extra
Mohawk	K/M	6400	KP	Back-light	80	\$8000	\$145	Pooler for 3 stations—\$175/month extra
Motorola	K/M	KB800	KP	None	200	\$8500	None	Pooler for 7 stations—purchase price \$9700
Potter	K/M	KDR	KP	BCD (Bit)	160	\$8100	\$165	Pooler for 3 stations—\$45/month
Sangamo	K/M	DS9100	KP	Back-light	120	\$8200	\$177	Pooler for 10 stations—\$247/month extra
Vanguard	K/M	Data-scribe	KP	None	200	\$8500	\$175	
Computer	K/T	Info	T	CRT	960	\$10,000 to	\$252 to	

Consoles Computer Entry System Mohawk	K/T	System 6000	KP	None	496	\$78,000 \$16,200 to \$42,000	\$200 \$360 to \$925	Two to 6 stations
	K/T	9000	KP	Back- light	80	\$53,000 to \$145,000	\$1040 to \$2840	Four to 16 stations
Computer Machinery	K/T	Key Process- ing	KP	Back- light	250	\$92,500 to \$168,100	\$2055 to \$4095	Eight to 32 stations
General Computer Systems Inforex	K/T	2100	T	Printed	200	\$81,240 to \$273,120	\$2350 to \$7885	Seven to 39 stations
	K/T	Key Entry	KP	CRT	128	\$30,300 to \$35,100	\$760 to \$960	Four to 8 stations
Penta Associates	K/T	Key Logic	KP	Back- light	200	\$110,000 to \$345,200	\$3000 to \$8600	Eight to 64 stations
Systems Eng.	K/T	Keytran	KP	None	300	\$100,000 to \$220,000	\$2875 to \$6350	Nine to 48 stations
Logic Corp.	K/D	LC-720	KP	CRT	350	\$148,000 to \$300,000	\$2450 to \$5800	Four to 16 stations

Legend:

K/T = Key to magnetic tape system

K/D = Key to disk system

K/C = Key to cassette

K/M = Key to computer compatible magnetic tape

KP = Key punch

T/KP = Typewriter or key punch

T = Typewriter

Backlight = a matrix consisting of all individual characters that can be keyed. Each character, as keyed, is displayed one at a time in its particular position in the matrix.

Projection and Light-emitting diodes = A one-character position dot matrix. Each character, as keyed, is displayed one at a time in the same position.

BCD (Bit) = Lights displaying the bit position (on, off) of individual characters. Each character, as keyed, is displayed one at a time.

(The prices quoted and the characteristics given of each device reflect the best information that could be obtained by the RECON staff.)

could be assigned to single keys and translated to their proper value by software, thus reducing the amount of keystroking required.

The Keymatic appears worth further investigation; therefore, the Library may rent a device for several months for testing and evaluation. A typist will be trained in current MARC/RECON procedures and assigned to the Keymatic as soon as her training period has been completed. The first month will be spent training on the Keymatic prior to the actual input of RECON records to obtain production and error rates and cost evaluation for comparison purposes.

Serious consideration was also given in the RECON Report to direct-read OCR equipment; however, at that time no equipment existed that offered the technical capability to perform the conversion of the LC record set. Since then, preliminary investigation of the Model 370 CompuScan Universal Optical Character Reader proved interesting enough to continue further exploration of the device.

The Model 370 CompuScan is a computer directed flying-spot scanner which matches the scanned portion of a character with a character described in the core memory of the computer.

The manufacturer has examined a sample of LC printed cards selected at random over a period of twenty years and has concluded that although the hardware is sufficient to read the record set optically, significant software effort would be required.

The results of the sampling indicated that the record set is not constituted entirely of "mint" cards, i.e., cards printed from the metal of the original Linotype composition, but is composed of originals and reprints of the original. When the stock of the original printing is close to depletion, the card is reprinted by photographing the card, and duplicates are made by a photo-offset process. As this cycle is repeated, the card for any one title could be several generations removed from the original. In some instances, a microscopic examination of the cards seems to indicate that the matrices used in the Linotype composition were worn. Because of these factors, what might appear as the same character to the naked eye would represent different pattern configurations to the scanner's core memory.

The coarseness of the card surface may also cause variations in the same characters. LC cards have a high rag content in order to meet the archival standards required by libraries. The roughness of the surface does not affect the readability for the human but may cause variations in a given character when read by an optical scanner.

Another significant problem with LC cards concerns characters which touch, i.e., connections between what are intended to be distinct characters but are read by the scanner as one. For example, if a lower case "n" were next to a lower case "t" and the cross bar on the "t" touched the "n," the scanner would consider the combination of the "n" and the "t" as one character.

Software must be written to handle the variant character and the touching

character problems. In the case of the touching characters, the machine must recognize some allowable limit of reading a single character, and when this limit is exceeded, the pattern read must be divided and matched against single-character patterns held in core. Programs can be written so that if either of the above conditions occurs, the output on magnetic tape will be flagged for later spot checking, permitting the scanner to continue to operate at throughput speeds without human intervention.

The resultant magnetic tape would serve as input to the Library's format recognition programs to reformat the scanner's output into the MARC II format. It has been estimated that the throughput speed of CompuScan would be in the vicinity of 1800 cards per hour.

The LC record set will be microfilmed according to the specifications required by the scanner. Since the scanner operates with negative film, a very dark background with a very clear, white image is necessary. A tentative cost estimate of the microfilming and reading has been computed at approximately fifty cents per 1000 characters output on magnetic tape (approximately three LC cards). This price does not include the cost of the software.

Original printed "mint" cards will be used to test the device without implementing the required software, and depending on the results, investigation may be continued.

The keying of the 1969 RECON records has been performed by a contractor using an IBM Selectric typewriter with the resulting hard copy fed through a Farrington optical character reader. As part of the contractor's services to the Library, production rates were monitored and reported. This gave LC the basis to compare two devices, the key-to-cassette used at the Library of Congress for the MARC Distribution Service and the equipment used by the contractor for RECON records.

To make the comparison in Table 5, it was necessary to determine the costs for each method using the techniques developed in the RECON report (9). Some modifications of cost were made to the original RECON estimates because actual figures are now available. MARC costs were obtained by dividing the costs of the manhours for typing and proofing in a given period by the number of records added to the MARC master file in the same period. The equipment cost per record was also based on the number of records added to the master file. Production rates associated with particular tasks were not used.

The manpower figures supplied by the contractor were limited to hourly production rates; therefore, to obtain the cost per record for OCR typing it was necessary to project the hourly rate to cover a manyear. The estimated annual production of a typist was then divided into the annual salary of a GS-4 (step 1) typist incremented by 8.5% for fringe benefits. The OCR equipment costs were computed on the basis of figures supplied by the contractor, assuming ownership of the OCR-font typewriter and service bureau rental of the scanner.

Table 5. *Input Costs per Record*1. *Manpower**Key to Cassette Method*

Typing	\$.45
Proofing	.70
	<hr/>
Total	\$1.15

OCR Method

Typing rate of contractor
 1,000 records in 104 hours or
 9.6 records per hour
 Typing cost at LC
 $\frac{\$5,522 + 8.5\% (\$5,522)}{9.6 \times 1,338} = \$.466$

Proofing rate of RECON editors at LC:
 1,534 records proofed in 173 hours or
 8.9 records per hour—
 20% = 7.1 records per hour
 Proofing cost at LC
 $\frac{\$6,882 + 8.5\% (\$6,882)}{7.1 \times 1,338} = \$.786$

Typing	\$.466
Proofing	.786
	<hr/>
Total	\$1.25

2. *Equipment* (costs do not include maintenance where applicable)*Key to Cassette*

Key to Cassette Monthly rental	\$100.00
Converter—Monthly rental prorated over 10 Key to Cassettes	26.00
Total	\$126.00

Hourly cost (assumes 132 hours a month) \$.955

Effective production rate of Key to Cassette

Average weekly MARC
 output
 $\frac{4 \text{ Key to Cassette units}}{120} = \frac{1,005}{120} = 8.4 \text{ records/hour}$

Record cost of Key to Cassette and converter

$\frac{\$.955}{8.4} = \$.114$

OCR Method

OCR-font typewriter		
Purchase price		\$500.00
40-month amortization		12.50/month
Hourly cost (assumes 132 hours use)		.095
Effective production rate of OCR typewriter		
$\frac{9.6 \text{ records/hour} \times 1,338 \text{ hours}}{132 \text{ hours} \times 12 \text{ months}} = 8.1 \text{ records/hour}$		
Record cost of OCR typewriter	$\frac{\$.095}{8.1} =$	\$.012
OCR scanner—service bureau hourly rental		\$ 50.00
10,000 lines/hour each record—		
18 lines 555 records/hour		
Record cost of OCR scanner		\$.09
Total record cost for equipment		
\$.012 + \$.09 =		\$.102

The cost of proofing in the OCR method was based on the RECON experience at LC modified by contractor experience. In actual practice, OCR records are proofed and corrected by the contractor before they are proofed by RECON editors. It was assumed that double proofing is unnecessary but that allowance should be made for the added difficulty of reading copy with a higher proportion of errors. (A preliminary study of errors on RECON proofsheets has shown that there are fewer typographical errors on RECON proofsheets than on current MARC proofsheets.) For this reason, the number of RECON records proofed in an hour has been decreased by 20% in the calculations.

On the basis of the calculations in Table 5, the comparative input costs are summarized as follows:

Table 6. Estimated Input Cost per Record

	<i>Key-to-Cassette</i>	<i>OCR</i>
Manpower: Typing	\$.45	\$.47
Proofing	.80	.78
Equipment	.11	.10
Totals	<u>\$1.26</u>	<u>\$1.35</u>

The final figures indicate that the two methods are very close in cost. As presently calculated, the key-to-cassette method is less expensive than the OCR method. It is easy to see that a slight change in any cost or production rate could make the OCR method less expensive. If the proofing

rate of 8.9 records per hour were maintained instead of decreasing to 7.1 per hour, the OCR proofing cost would drop to \$.63, and the total price for this proposed method would be \$1.20.

One way to test the assumption of the added difficulty of a single proofing would be to obtain uncorrected records from the contractor as a means of determining the actual proofing rate under that condition.

RECON Tasks

The four tasks that have been identified for study by the Working Task Force are: 1) levels of completeness of MARC records; 2) implications of a national union catalog in machine readable form; 3) conversion of existing data bases in machine readable form for use in a national bibliographic service; and 4) study of problems involved in any future distribution of name and subject cross reference control files. Progress to date on the first three tasks is described in the following paragraphs.

Task 1 has been completed, and an article summarizing the results of a report submitted to CLR has been published in the *Journal of Library Automation*, June 1970 (10). The following conclusions reached by this study are quoted from the article: 1) The level of a record must be adequate for the purposes it will serve. 2) In terms of national use, a machine readable record may function as a means of distributing cataloging information and as a means of reporting holdings to a national union catalog. 3) To satisfy the needs of diverse installations and applications, records for general distribution should be in the full MARC II format. 4) Records that satisfy the NUC function are not necessarily identical with those that satisfy the distribution function. 5) It is feasible to define the characteristics of a machine readable NUC report at a lower level than the full MARC II format.

Task 2 consists of an investigation of the implications of a national union catalog in machine readable form. A design of such a system is needed, and although the implementation of such a project is beyond the purview of the Working Task Force, some of the technical and cost factors should be examined and defined for possible future research.

As a framework for discussion purposes, a future reporting system for the National Union Catalog was postulated based on the present reporting system as follows:

<i>Contributors</i>	<i>Present Report Form</i>	<i>Future Report Form</i>
LC	Printed cards	LC - MARC data (for all records)
Outside libraries	Locally produced cards and LC cards	MARC data (for all records) or records submitted to NUC to be keyed as machine readable records

The problems of the control number and library location symbols were considered, but a tentative decision was made that recommendations should be forthcoming when the American National Standards Institute Sectional Committee Z39 has completed its work on library identification codes. The indicators and subfield codes to be included in the machine readable NUC records would depend on the optimum file arrangement of the suggested bibliographic listings. The Library of Congress is presently engaged in a filing rules study which should influence the inclusion or exclusion of particular content designators. Task 2 is still in progress.

Task 3 is the investigation of the possible utilization of other machine readable data bases for use in a national bibliographic store. The task was divided into several subtasks as follows: 1) identification of useful data bases for the purposes described (content and bibliographic completeness); 2) cost of the conversion from a local format to a MARC II record; 3) cost of updating records not already in the LC data base for consistency and missing data by comparing the records with the Library of Congress Official Catalog; 4) cost of comparing the record for the existing LC machine readable records to eliminate duplicate records. To satisfy the first subtask, a questionnaire was sent to 42 organizations. The information requested included:

- 1) Availability of data bases—maintained by library or service bureau, and permission to copy data base.
- 2) Use of the data base—for acquisitions, production of book catalog, circulation system, etc.
- 3) Composition of data base—monographs, serials, technical reports, etc.
- 4) Composition of data base—number of titles, imprint dates (primarily current, retrospective, etc.), language of records.
- 5) Source of catalog data—MARC Distribution Service, LC catalog card, local cataloging.
- 6) Data elements for monographs.
- 7) Format used in identifying data elements—MARC I format, MARC II format, etc.
- 8) Character set used.

The results from this survey were analyzed, and a follow-up letter was sent to 22 of the organizations, requesting further information as follows:

- 1) An estimate of the number of monographs added to the data base each year.
- 2) Representative group of twenty-five entries for monographs including both fiction and non-fiction.
- 3) Details on the character set used in the machine readable data base.
- 4) Detailed specifications of monographic record format.

Responses from this last letter have been received and analyzed. This analysis should identify a limited number of machine readable data bases that will be subjected to further content and cost analysis.

OUTLOOK

The RECON Project continues to be on schedule. The Working Task Force has met several times for deliberations on the assigned tasks; in addition, members have been briefed on the progress of the pilot project and their advice has been sought. Thus, individuals interested in the problems of bibliographic conversion guide the project throughout its development.

The Library of Congress RECON staff continues to maintain liaison with individuals and organizations working in any facet of the project's scope, hoping to bring all expertise possible to bear on the problems involved.

It is significant, although not fully recognized at the onset of the RECON Project, that the solution to many of the problems under exploration will have impact on current conversion as well as retrospective conversion. This is evident at the Library of Congress where MARC and RECON, although staffed separately in the production area, share staff in the Information Systems Office, and the project is known as MARC/RECON.

Coordination continues between the RECON Project and the Card Division Mechanization Project. The RECON Project Director is the technical adviser for the Card Division Project, and under her general direction, a computer analyst in the Information Systems Office has been assigned full time to the project. The analyst has been given a detailed orientation to the procedures and computer programs for MARC/RECON and the specifications for the Card Division Project. This exposure is necessary to guarantee that there is no duplication of effort between the two projects and that the design work for the Card Division Project includes the possibility of a future national service for machine readable cataloging, both current and retrospective. (The MARC Distribution Service is such a national service for English language monograph cataloging data, but what is assumed here is a service of a much broader scope.)

Although progress has been made in many of the tasks included in RECON, several methods of input described in the RECON Report can only be fully evaluated when the format recognition programs are implemented. According to present estimates, this should take place toward the end of 1970.

Much remains to be accomplished. The Library of Congress will continue to make its progress known as rapidly as possible, because the results of the pilot project will have great ramifications for the entire library community.

ACKNOWLEDGMENTS

The authors wish to thank the staff members associated with the RECON Pilot Project in the Technical Processes Research Office and the MARC Editorial Office in the Library of Congress Processing Department, and

those in the Information Systems Office, for their respective reports, which were incorporated into the progress report submitted to the Council on Library Resources and which provided significant contributions to this paper.

REFERENCES

1. Avram, Henriette D.: "The RECON Pilot Project: A Progress Report," *Journal of Library Automation*, 3 (June 1970).
2. RECON Working Task Force: *Conversion of Retrospective Records to Machine-Readable Form* (Washington, D. C.: Library of Congress, 1969), pp. 32-33.
3. Avram, Henriette D., et al.: "MARC Program Research and Development: A Progress Report," *Journal of Library Automation*, 2 (December 1969), 250-253.
4. RECON Working Task Force: Op. cit., p. 31.
5. National Microfilm Association: *Glossary of Terms for Microphotography and Reproductions made from Micro-Images*. 4th rev. ed. (Annapolis, Md.: National Microfilm Association, 1966), p. 8.
6. Ibid.
7. Ibid., p. 52
8. Hawken, William R.: *Copying Methods Manual* (Chicago: Library Technology Program, American Library Association, 1966), p. 243.
9. RECON Working Task Force: Op. cit., pp. 58-59, 86, 93.
10. RECON Working Task Force: "Levels of Machine-Readable Records," *Journal of Library Automation*, 3 (June 1970).