LIBRARY CATALOGUE AUTOMATION: COST-BENEFIT FACTORS

Fred W. Matthews
School of Library Service
Dalhousie University
Halifax, N.S.
B3H 4H8

ABSTRACT

The cost of operating a catalogue system based on "MARC" type records and one using selected fixed length fields is compared. The factors which contribute to the approximately ten times lower cost of the selected system are outlined.

AUTOMATISATION DES CATALOGUES DES BIBLIOTHEQUES: FACTEURS: FRAIS - AVANTAGES

RESUME

Cette étude compare le coût d'opération d'un système basé sur les enregistrements (relevés) "MARC" et un système utilisant un champ d'accès fixe et choisis. Nous soulignons les facteurs contribuant au résultat d'un coût approximativement dix fois moins élevés du système choisis.

INTRODUCTION

The replacement of the card catalogue with an automated book or fiche catalogue is proceeding rapidly in National, University and Research libraries. The announcement that the Library of Congress is planning to close its card catalogue gives all concerned with catalogue maintenance reason to think to themselves "if they can do it, why can't I?" The imminent adoption of AACR II, the new Anglo American cataloguing rules, brings both a convenient time to close the present catalogue and a sense of urgency to making a move. While much information is available for these large projects, relatively little is available to assist the smaller libraries in the design of a system for their particular needs. These smaller libraries include public, school and undergraduate libraries, of which there are many more than of large libraries. The "benefits" which accrue to the large libraries would be similar in nature and would be multiplied many fold if automation is successfully extended to include these smaller libraries. A system has been developed for the Dartmouth Regional Library (1) (6) Dartmouth, Nova Scotia which for the purposes of this paper serves as a prototype for a system meeting the needs of a public library network with a collection size of about 100,000 items.

THE MARC RECORD

The Machine Readable Catalogue (MARC) system first developed at the Library of Congress, Washington, (3) has become the prototype for the system for international transfer of full bibliographic descriptions. To fulfill this role, provision has been make for the inclusion of data of many possible types, each precisely designated to meet the needs of human input and machine readability and processing. A comprehensive manual is required to explain the intricacies of the record. The system uses many data fields which may contain a variable number of characters, letters, numbers or symbols - i.e. variable length fields. Within each field there is a possible set of sub-fields designated by indicators, the meaning of which changes for each field. The field tags are a three digit number,

allowing for 999 types.

While Author, Title, Imprint and Subject fields occur in most records, the system provides for specialized tags to differentiate additional types of data needed for a full bibliographic description. As the system has been extended to cover maps, music, recordings, and films, many more tags have been assigned specialized roles in the system. The guide to MARC tags issued by the Library of Congress (4) lists some two hundred tags with the indicators which apply to each. A basic description of the system is given in the text book by Hayes and Becker (2).

In planning for as efficient as possible data processing of such a variable and complex a record, the system requires in addition to the data itself, special elements, eg. a Record Leader of 24 characters which gives basic information on the record. In addition a Directory is required which is in effect a "table of contents". This lists all tags present in the record their length and starting position. Having located by means of the tag a specific type of data in the directory, the machine can locate these data by address and the character count given in the Directory. Character counting is a very fast and efficient operation for the computer. The logic required for the identification of sub-field indicators in a variable length field of characters is, however, a much slower operation. As all variable fields have sub-field indicators no data can be moved or manipulated without the symbol identification which for each character by character moved, in effect, asks the question "is this an indicator?". If the answer is yes, then something new must be done. This step increases very markedly the time required to manipulate MARC type records in the computer.

While MARC records are required for international transfer of bibliographic data, their use as the base record in a given situation should depend on the needs of the institutions concerned, in particular their users needs. National, Research and large University libraries need the completeness and flexability provided by the MARC format, for public, school and many others, a shortened content and simplified format can be considered.

In Australia the University of New England in New South Wales and James Cooke University has developed records of restricted content as a means of reducing cost of catalogue maintenance.

A SIMPLIFIED RECORD

In designing a record for the Dartmouth Regional Library system, the needs of the users was primary, however, efficient use of the computer was also considered. The staff at the Dartmouth Library were willing to take a practical approach which they felt was workable without insisting on the niceties of cataloguing procedures. The fact that the catalogue system has been used for four years without evident difficulties due to content or format, virtually without change, shows how effectively these initial decisions were made. Table I shows the content of the record and Fig. I the page format. The system uses fixed length - 127 characters and fixed fields. For each item there is a basic or primary record and provision for up to eight subject, two added authors and three added titles or series

entries. Any item could have a maximum of nine records in a set. In practice the average number of records per item is 1.2 (85,000 records for 70,000 items) which includes the cross reference entries which have been added to the subject catalogue. The figure is lower than might be expected as about 25% of the collection is fiction which normally requires only a single record. A single record can carry two subject headings which meets the needs of many items. By contrast items such as automobile manuals and phonodisks normally require two additional records. An anthology may require the maximum set of three author and four title records, or double this if a second set is used.

The imprint information is not included in the catalogue except for the last two digits of the year of publication. This is typical of the kind of compromise that was made in designing the record. The library maintains a shelf list card file which carries imprint, publisher and date as well as the number of copies ordered, date ordered, price, etc. This information is held primarily for internal use, and is not normally a major concern of the user of the library. The year of publication was however considered of sufficient interest that it should be included in perhaps an abbreviated form. Since very few books in the library were published before the year 1900, it was decided to use only two characters to designate the year of publication - with 00 used for 1900 and before. A computer program to count the entries in the catalogue (70,000 entries) by imprint year showed that there are 84 books in the collection published before 1901. This kind of statistical information on the collection shows the type of data, easily available, from a machine readable catalogue file. The distribution curve of the number of books for each imprint year forms a profile of the collection that is useful on its own or for comparison with collections of other libraries. This aspect will be the subject of a future paper.

The Author field is limited to 29 characters. This presents no difficulty for personal authors which, even with forenames included in full, rarely exceed twenty-nine characters. In the case of corporate authors, an authority file is maintained to give uniformity of entry for this class of author. The title subject field is eighty characters. The title and two subject headings may appear in this area. Each of the subject entries is preceded by an "=" sign which designates the subject headings for both the user and the computer system. Three added subject records carry two subject entries each making eight possible subject entries. In the Dartmouth library, Sears subject headings are used. These are less detailed and are shorter, on the average, than Library of Congress Subject Headings (LCSH) and often allow the title and two subject headings in the prime record. With long subject headings, the number of subjects may be reduced to one per subject record for a total of four. Very few items (150 in 71,000) have used more than four subject entries.

The one character indicators in the record are identified in Table I. In addition to providing location information in the catalogue, these allow specialized catalogues to be produced. To date, selected catalogues have been produced using as selection criteria: Large Print,

French language, Talking Books and Easy Read. These short catalogues can be printed in a sufficient number of copies to allow for wide distribution, e.g. in the case of Large Print to Senior Citizens' Homes. Selected Catalogues based on subjects and/or authors can be produced by the same routines. - See Fig. II.

TABLE I

CONTENT OF THE DARTMOUTH CATALOGUE RECORD

Collection Indicator	- one character
	J - childrens
	D - Dartmouth Municipal Documents
Type of Item	- one character
	F - Fiction
	R - Reference
	E - Easy Read-Low Vocabulary
Dewey Number	- 8 characters
	Class number - four digits after the decimal point.
Format Indicator	- one character (22 indicators)
e.g.	B - Talking Books
	D - Phonograph Records
	F - Books in French
	L - Large Print Books
	X - Oversize Books
Imprint Year	2 characters
Author	29 characters
Title/Subject	80 characters
Added Subjects	60 characters
Added Authors	29 characters
Added Title or Added Series Entries	80 characters

SHORT ROUTINES

Much of the value of automation of catalogue production results from the speed, accuracy and low cost of machine sorting. In card catalogue maintenance this cost is somewhat obscured by the nature of the work and is revealed only if a detailed study is carried out. In fixed field records the sort fields, i.e. authors in the author catalogue, Dewey numbers in the shelf list catalogue, etc. are easily defined. In the MARC record provision has been made using an indicator, for the number of lead characters to be ignored in sorting for "The" at the start of a title ignore four characters. This necessitates special handling of each record before the "sort field" can be designated and adds to the cost of sorting type records.

PAGE FORMAT

Whether using Computer on Microfilm (COM) fiche or film catalogues or printed book catalogue, the "page" format for presentation of the catalogue deserves careful consideration. Two factors are most important, the ease of scanning entries and the number of entries on a page or frame. Much of the use of catalogues concerns scanning entries so this factor becomes of prime importance. The Dartmouth catalogue uses a one line entry with fixed fields and fixed line format. This provides for a very effective scan of the catalogue and fifty entries per page or COM frame. (Fig. I) Catalogues derived from MARC records frequently revert to the cardlike format from which they were derived. This often gives only about fifteen entries per page and relatively poor scan characteristics. Such as they are, the scannable characteristics are derived only after extensive moving about of data to get extended margin or indented entries. This is repeated for each of the catalogues produced. This added data movement adds to the cost of manipulating MARC data in the computer. By contrast the one line catalogue uses the same line format for three of the four catalogues. An interchange of subject and title is only data movement. This is required for the subject catalogue.

The lower number of entries per page will add to the number of pages or frames in the fiche catalogue. An increase of three times in the number of pages will be a major factor in a printed book catalogue and, while not so evident in a fiche catalogue, the cost is increased and convenience of searching the catalogue is decreased as the number of fiche grows.

COSTS

It is difficult to get comparative cost comparison of catalogue systems. For a public library operating from a MARC based system, the data reported here refers to the Public Library in Missassagua, Ontario which used the University of Toronto Library Automated System to derive a book catalogue. The costs were given at a Canadian Library Association Seminar in Halifax as $2.38 per title. The factors entering into this cost were not given. The Dartmouth cost is given (6) as 20¢ per title. This includes preparation of the machine readable input, computer processing of the data and printing the required author, title and subject catalogues in nine copies, shelf list in Dewey Class Number order, (3 copies) and Master file listing (1 copy). The latter is used for control of deletions and corrections to the records. The cost of both the Annual full catalogue and the cumulative monthly additions catalogues, and monthly accession lists are included. Further details on the cost are given in the paper published in the Canadian Library Journal (6). Table 2 give costs on printing the 1979 Annual Catalogues. Based on the information given in this table the cost of operation of the catalogue is $1400.00 per year. This could be prorated on the 13,000 new records added and would give a cost of about $0.11/record added or $0.13¢/title added. If prorated on the 78,000 records in the system it would give annual maintenance cost for this

catalogue for each item in the system of about 1.9¢/title. The charges for preparing the input takes at the City Hall in Dartmouth are not included as no charges are made. It is estimated that the annual cost of this service would be about $200.00/year. Since preparing input using a terminal is lower than the previous method of using input forms and punching cards, the costs are now lower than previously reported.

The cost does not include the assigning of Dewey numbers or subject headings. These intellectual functions, can add major amounts to the cost of preparing cataloguing entries for any system. The use of derived cataloguing i.e. using the cataloguing data prepared by other institutions is a clear way of reducing the intellectual costs of cataloguing.

Table II

Dalhousie Computer Charges - March/78-February/79

		Charges
Monthly Catalogues - 9 runs, Cumulative to 13,018 records		
each 9 copies Author Catalogues 9 copies Title Catalogues 9 copies Subject Catalogues 3 copies Dewey Catalogues		$877.03
Annual Catalogues		
Merge of Master Files		
Cumulative Monthly - 1978-	13,018 records	
Annual - 1978	71,382 records	22.48
Annual Update		
Delete. 6,689 records	78,711 records, net	52.00
Annual Create Records		
Create and Sent. Dewey Catalogue	51,856 records	
Author Catalogue	75,503 records	215.86
Title Catalogue	72,635 records	
Subject Catalogue	74,318 records	
Annual Print		
Print 3 copies of each of 4 catalogues		68.00
Print 9 copies of each of 3 catalogues		162.00
		520.34
Total Annual Computer Charges		$1397.37

* Normal Commercial Rates on the Dalhousie CDC 6400 computer

Approximate	$400/hour CPU Time
Print Charge	50¢/1000 lines

DERIVED CATALOGUING

The MARC records prepared by the National Library of Canada, Great Britain, the United States and Australia provide, for books published in English, cataloguing information of high quality. The MARC Records Distribution Service of the National Library of Canada (NLC) supplies these records from their comprehensive MARC databank in machine readable form, on magnetic tape, or on cards. The cost is about four cents a record for tape records and seven cents for a card. Using International Standard Book Numbers (ISBN) as the identifier, records for about 85% of the requests were returned in the four trials of the service involving six to seven hundred items.

A program to create Dartmouth type records from MARC records has been written. While satisfactory for a monthly accessions Bulletin, about half of these automatically prepared entries require editing before permanent addition to the catalogue. The main difficulty encountered concerns the Library of Congress Subject Headings (LCSH) which are more detailed than Sears Subject Headings and less suitable for a collection of a public library of the size of 100,000 items. If the library used LCSH less difficulty of this type would be encountered.

COST COMPARISON

A number of factors have been discussed which affect the cost of automation of a catalogue of books. The marked difference quoted for a MARC Based book catalogue and the Dartmouth Catalogue calls for some explanation. The content of the record is certainly a major factor. If data is carried in the catalogue files that is not required by most users, a cost penalty is imposed on the system. The use of fixed length records is probably the most important factor in simplifying the processing of the data. While this incurs some penalty in repeating data, and leaving blanks in the records, the penalty for these factors is small compared with the costs of manipulating a complex record of the MARC type. While fixed length fields present potential difficulties in "fitting" data into the space allowed; this has presented virtually no problems in the 70,000 records prepared for the Dartmouth Library. Decisions at this level are made by the technicians entering the data.

Sort routines add considerably to the computer costs when large files have to be handled. Page format can add substantially to the cost and bulk of the final product. The easy use of the resulting page deserves close attention. While these appear to be the major factors, further work is in progress to identify further factors and to determine the contribution of each to lowering the costs.

CONCLUSION

In designing a system the needs of the user population are paramount. No attempt should be made to satisfy all the needs of all the users but rather to satisfy most needs at a cost commensurate with value. If a

benefit can be provided e.g. publication year, at low cost, the addition should certainly be made. The use of data derived from existing sources of quality data can reduce the costs of both the intellectual and the mechanical input. Since this works two ways it can be considered a major factor in implementing and maintaining an automated system. When the costs of repetitive cataloguing in many locations is considered, the alternative of this type of network resource sharing is basic to the cost sharing of catalogue maintenance.

ACKNOWLEDGMENTS

In developing the system at the Dartmouth Regional Library the support of the project by the Chife Librarian, Heather Harbord and later, Aileen Barker was essential. In the specification of content of records, sort fields and page format, Lynn Murphy, Barbara Shuh and Susan McLean were most helpful. In the design of the computer system, Sheldon Hartling and Leslie Foster were of major assistance. In the everyday running of the system, Emile Theriault on the staff of the Dartmouth Regional Library has shown the care and attention to detail, without which no system will work properly.

REFERENCES

(1) Harbord, Heather. The Computer Catalogue at the Dartmouth Regional Library. APLA Bulletin, vol. 41, no. 2, page 31 (1977).

(2) Hayes, R.M., Becker, J. Handbook of Data Processing for Libraries. 2nd Edit. p. 471. (1974).

(3) Library of Congress. The MARC II Format. Washington, D.C., U.S.A. (1968).

(4) Library of Congress. MARC Development Office Composite MARC Format. Washington, D.C., U.S.A. (1974).

(5) Matthews, Fred. W. Library Catalogue Automation for a Public Library. Proceedings of the Canadian Society for Information Science,(1976). page 117.

(6) Matthews, Fred W. Library Catalogue Automation Optimizing the Cost. Canadian Library Journal (in press).

DARTMOUTH REGIONAL LIBRARY — PAGE 3

DEWEY	YEAR	AUTHOR	TITLE AND SUBJECT(S)	
J 398.	F75	ANDERSEN, HANS CHRISTIAN	PETITE FILLE AUX ALLUMETTES =FAIRY TALES	A07641
663.	76	ANDERSON, [illegible] SEE GEORGE, J	EASY WINE MAKING IN TWENTY ONE DAYS	A03945
697.	X76	ANDERSON, BRUCE M	SOLAR HOME BOOK =SOLAR ENERGY	604271
597.	X76	ANDERSON, GARY J	SOLAR,STORY OF THE ATLANTIC SALMON	604411
971.	76	ANDERSON, GRACE M	FUTURE TO INHERIT =CANADA - HISTORY =PORTUGUESE IN CANADA	A07551
J 398.	F72	ANDERSON, HANS CHRISTIAN	COMTES DANDERSEN =FAIRY TALES	605051
746.	X71	ANDES, EUGENE	PRACTICAL MACRAME =MACRAME	605601
301.36	75	ANDREWS, GEORGE F	MAYA CITIES =MAYAS - ARCHITECTURE =MEXICO - ANTIQUITIES	A04091
JFICT.	77	ANGIER, BRADFORD	ASK FOR LOVE AND THEY GIVE YOU RICE PUDDING =FAMILY LIFE - FICTION	A04041
P581.6	76	ANGIER, BRADFORD	COLOR FIELD GUIDE TO COMMON WILD EDIBLES =PLANTS,EDIBLE	599391
FICT.	76	APPLE, MAY	ORANGING OF AMERICA	604761
JFICT.	68	ARDEN, WILLIAM	ALFRED HITCHCOCK AND THE THREE INVESTIGATORS IN THE MYSTERY OF MOANING CAVE	A 03941
635.966X72		ARGYLE, CHARLOTTE	PRESERVING NATIVE FLOWERS,HOW TO DISPLAY THEM =FLOWERS - DRYING	605121
796.	75	ARLOTT, JOHN, ED SEE	OXFORD COMPANION TO WORLD SPORTS	605665
R796.5	77	ATKINSON, G SEE JENGTSSON, H	ORIENTEERING FOR SPORT AND PLEASURE	605665
971.6	ND	ATLAS DE LACADIE	PETIT ATLAS DES FRANCOPHONES DES MARITIMES =ACADIA	A04051
921.	F71	AUBERTDEGASPE, PHILIPPE	MEMOIRS =AUBERT DE GASPE, PHILIPPE	A04041
J 398.	68	AUBREY, CLAUDE	MAGIC FIDDLER =FOLKLORE,FRENCH CANADIAN	604601
793.32	X76	AUDY, ROBERT	ROBERT AUDY METHOD TAP DANCING =DANCING	605351
FICT.	77	AUSTEN, JANE	WATSONS	607671
JFICT.	77	BACH, ALICE	FATHER EVERY FEW YEARS =SINGLE PARENT FAMILY - FICTION	AL 9481
784.	X70	BACHARACH, BURT F	BACHARACH AND DAVID SONG BOOK =SONGBOOKS	605521
FICT.	76	BAINBRIDGE, BERYL	QUIET LIFE	606751

DARTMOUTH REGIONAL LIBRARY — PAGE 2

DEWEY	YEAR	AUTHOR	TITLE AND SUBJECT(S)	
J 599.8	X76	COOK, DAVID	=APES >CLOSER LOOK AT APES	610121
616.8	77	DAHLBERG, CHARLES CLAY	=APHASIA >STROKE	599725
923.1	76	COREN, LEONARD	=APPLICATIONS FOR POSITIONS >GET THE JOB YOU WANT	605431
P331.1	75	EPPLEN, JOHN D	=APPLICATIONS FOR POSITIONS >JOB HUNTING GUIDE	605171
913.031	77	MACAULAY, ROSE	=ARCHEOLOGY >PROUDLY BENT INTERPRETS IN PHOTOGRAPHS PLEASURE OF RUINS	605161
799.32	54	BURKE, EDMUND H	=ARCHERY =ARCHERY, HANDBOOK	605141
399.	X72	TUNIS, EDWIN	=ARMS AND ARMOR >WEAPONS,A PICTORIAL HISTORY	A03961
246.	X75	BRANSON, SAMUEL GEORGE	=ART AND RELIGION >MAN AND GOD IN ART AND RITUAL	605451
709.6	X74	DUERDEN, DENNIS	=ART,AFRICAN >AFRICAN ART	603721
R000.	ND	Z3 NEWSLETTER	=ARTHRITIS >C.A.R. SCOPE ASK AT DESK	YA6 21
759.11	76	FORRESTALL, TOM	=ARTISTS,CANADIAN >THIS GOOD LOOKING LAND	604821
759.4	X49	RUSSELL, JOHN	=ARTISTS,FRENCH >WORLD OF MATISSE,EIGHTEEN SIXTY NINE TO 1954	604541
R000.	ND	Z3 MAGAZINES	=ATLANTIC PROVINCES >ATLANTIC ISSUES - KEPT INDEFINITELY	573001
921.	F71	AUBERTDEGASPE, PHILIPPE	=AUBERT DE GASPE, PHILIPPE >MEMOIRS	A04141
621.38	76		=AUDIOVISUAL EQUIPMENT >AUDIOVISUAL MARKET PLACE,A MULTIMEDIA GUIDE	603861
JFICT.	74	CLARK, MAVIS	=AUSTRALIA - FICTION >SKY IS FREE	604321
JFICT.	77	SOUTHALL, IVAN	=AUSTRALIA - FICTION >WHAT ABOUT TOMORROW	604321
R920.	76		=AUTHORS - BIOGRAPHY >INTERNATIONAL AUTHORS AND WRITERS WHOS WHO, 7TH ED	573461
029.	74	MEREDITH, SCOTT	=AUTHORSHIP >WRITING TO SELL, 2D REV	A 4131
808.	73	TUPABIAN, KATE,LARIMORE	=AUTHORSHIP - HANDBOOKS,MANUALS,ETC >MANUAL FOR WRITERS	604302
808.	75		=AUTHORSHIP - HANDBOOKS, MANUALS,ETC >BETTER LETTERS AND REPORTS	603483
808.	69	TUPABIAN, KATE L	=AUTHORSHIP - HANDBOOKS,MANUALS,ETC >STUDENTS GUIDE	603382
629.2	76	EDMONSTON, PHIL	=AUTOMOBILES >LEMON AID, 1977 EDITION	597171

Fig. I - Author and Subject Catalogues.

The author, title and Dewey catalogues have the same line and page format. Only the order on the page is changed. For the Subject Catalogue, The Subject is moved before the title providing a sortable field for the subject entries. The = sign indicates a subject entry, the > the title.

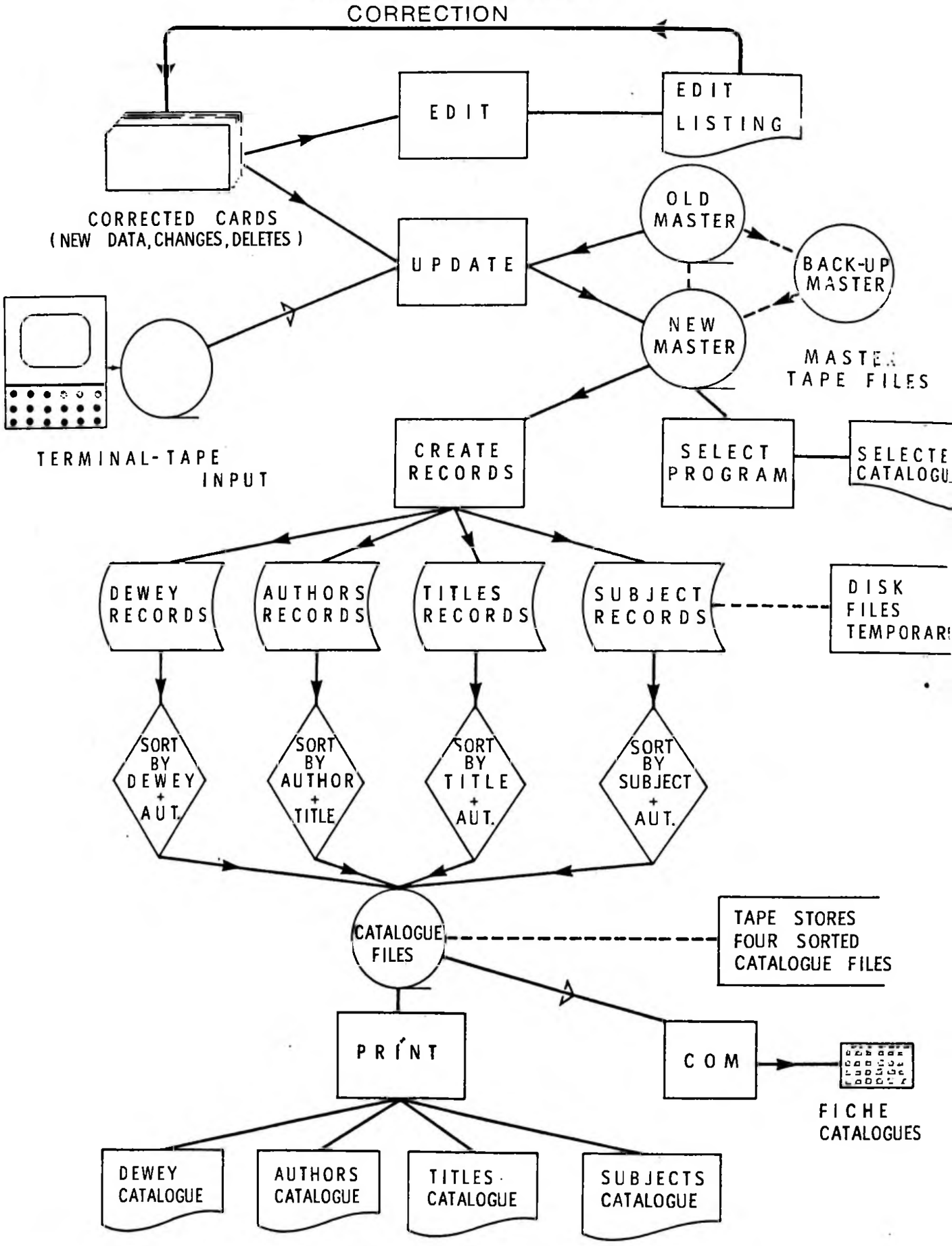

FIG. II Systems Flowchart

After the EDIT phase cards or tape input go to the UPDATE to create a New Master File when new catalogues are required, this file goes to CREATE RECORDS where four separate catalogues are prepared which when sorted, become Files. These may be sent to PRINT or COM to produce Book or Fiche Catalogues.