

# The Minnesota Union List of Serials

Audrey N. GROSCH: University of Minnesota Libraries

*This paper describes development of a MARC serials format union catalog of serials called the Minnesota Union List of Serials (MULS). The Preliminary Edition, published August 1972, contains over 37,000 main entries in 1,566 text pages produced through photocomposition in News Gothic typefont using the full MARC character set. The total number of entries is over 59,000, including cross-references. Conceptualization and scope of the system as well as its design, data conversion, computer and programming support, photocomposition, costs, and problems are discussed.*

## INTRODUCTION

This paper has been prepared to inform the profession of the development of the Minnesota Union List of Serials (MULS), the data base of which represents significant differences from those of previously reported union lists. As one can see from Figure 1, a MULS Preliminary Edition sample page, MULS is a full bibliographic union serials catalog. It uses the MARC serials format as its formal structure. The Preliminary Edition contained 37,289 University of Minnesota serial titles. The file now includes holdings of the Minneapolis Public Library, eight private colleges, and ten Minnesota State Agency libraries including the Minnesota Historical Society. Augmentation of the data is continuing so that all Minnesota academic, large public, and selected special libraries will be included in the coming two years.

Several years ago, the University of Minnesota began investigating the development of its own unified serials catalog as a first stage in the development of an automated serials management system. At that time many libraries in the state had developed their own serials lists, and regional consortia had created lists of their members' holdings. These resources, coupled with networking through the MINITEX (Minnesota Inter-Library Teletype Exchange) program, made possible the MULS initial development. The MINITEX program links together seventy-two libraries via teletype to the University of Minnesota for rapid interchange of library materials. State supported academic institutions, public libraries, and private colleges in the local metropolitan area also participate in this program.

In spring 1971, it became apparent that a union list had become a necessity if the expanding MINITEX program and the university were to

## Minnesota Union List of Serials

Franciscan historical

France. Ministère de l'Air. Publications scientifiques et techniques. Bulletin des services techniques. Paris. MnU En No 1. (1921- ).	2073800	France. Office de la Recherche Scientifique et Technique outre mer. Bulletin Analytique d'entomologie medicale et veterinaire. Paris. Title varies: volumes 1-16. Bulletin signalétique d'entomologie medicale et veterinaire. MnU A ENT V.1:2,3 (1953- ).	2075100
France. Ministère de l'éducation nationale. Annuaire de l'éducation nationale. Paris. "Publié par l'Institut pédagogique nationale." MnU WIL 1962,1965,1968,1970-370.58 of 5844 MnU R Current volume only.	2073900	France. Office de la Recherche Scientifique et Technique Outre-Mer. Bulletin signalétique d'entomologie medicale et veterinaire. Paris. Title varies: Bulletin analytique d'entomologie medicale et veterinaire. MnU A VET V.14. (1967- ).	2075200
France. Ministère de l'éducation nationale. Catalogue des theses de doctorat. Paris. At head of title: 1884/1885-1931: Ministère de l'instruction publique; 1932-1958 ministère de l'éducation nationale, 1960-1968, 1884/89-1958: Issued annually, 5 yearly parts being pagged continuously and supplie d with a general I, p. and subject index to form a volume. Vols. 1-5, 8 (1884/89-1904/09, 1919/23) have also an alphabetical list of authors. In vols. 6-7 (1909/13-1914/18) a list of authors is given with each fascicle. Beginning with 1959, vol. numbers no longer used. Beginning with theses of 1930, a list was also published in parts, as supplements to the numbers of the Bibliographie de la France, 1. p. 1e, Bibliographie, 1932 date. Title varies: 1884/89-1958, Catalogue des theses et écrits academiques. 1965- : called Bibliographie de la France, Supplément Annuel D 1915 16 (fasc. 32-33) and 1917 18 (fasc. 34-35) each issued in one number. MnU BIB V.1-15 (1884/1889-1958); n. ser. 1959-016.378 of 84	2074000	France. Office Scientifique et Technique des Pêches Maritimes. Revue des travaux. Paris. MnU A ENT V.1:11 (1928-1938).	2075300
France. Ministère de l'instruction publique et des beaux-arts. See France. Ministère de l'éducation nationale. Catalogue des theses de doctorat.	2074100	France. Office nationale de la navigation. Statistique Annuelle de la Navigation Interieure. Paris. Title varies slightly. Supersedes in part its statistique de la navigation interieure; nomenclature et conditions de navigabilité des fleuves, rivières et canaux, relevé general du tonnage des marchandises. MnU Doc 1969-56 34826	2075400
France. Ministère de l'instruction publique et des beaux-arts. See Bulletin des bibliothèques et des archives.	2074200	France. Office Scientifique et Technique des Pêches Maritimes. Revue des travaux. Paris. MnU A ENT V.1:11 (1928-1938).	2075600
France. Ministère de l'instruction publique et des beaux-arts. Comité des travaux historiques et scientifiques. See France. Comité des travaux historiques et scientifiques. Section de Géographie. Bulletin.	2074300	France. Secrétariat de l'Etat à l'aviation. Publications scientifiques et techniques. See France. Ministère de l'air. Publications scientifiques et techniques.	2075700
France. Ministère de l'instruction publique et des beaux-arts. Comité des travaux historiques et scientifiques. See France. Comité des travaux historiques et scientifiques. Section de Géographie. Bulletin.	2074400	France. Service des études économiques et financières. Rapport sur les comptes de la nation. Paris, Impr. nationale. Vol. for 1949-1955 issued by service des études économiques et financières and Institut national de la statistique et des études économiques. 1962-1966 by Institut national de la statistique et des études économiques. 65 42400 MnU Doc 1967-336 44 of 8474 MnU WIL 1960.1966.	2075800
France. Ministère de l'intérieur. Revue generale d'administration. Paris, Impr. Berger-Levrault. At head of title, 1879-1910: Ministère de l'intérieur. Title varies slightly. V. 1-3, Jan. Dec. 1878. Publiée avec la collaboration de sénateurs, de députés, de membres du Conseil d'Etat, de fonctionnaires et de publicistes sous la direction de Maurice Block. 8 4538 rev MnU PER 1878-1910.	2074500	France. Service des Pêches Maritimes. See France. Office Scientifique et Technique des Pêches Maritimes. Revue des travaux.	2075900
France. Ministère des Affaires Etrangères. Commission de Publication des Documents Relatifs aux Origines de la France. Commission de Publication des Documents Relatifs aux Origines de la Guerre de 1914. Documents diplomatiques...	2074600	France. Services techniques et industriels de l'aéronautique. Publications scientifiques et techniques. See France. Ministère de l'air. Publications scientifiques et techniques.	2076000
France. Ministère des finances. See Bulletin de statistique et de législation comparée.	2074700	France. Armée. Etat-major. Service historique. See Revue historique de l'armée: publication trimestrielle de l'Etat-major de l'Armée. Service historique.	2076100
France. Ministère des finances et des affaires économiques. Rapport sur les comptes de la nation. See France. Service des études économiques et financières. Rapport sur les comptes de la nation.	2074800	France actuelle. Paris. MnU Doc V.15 no.1- (JAN1966- ).	2076200
France. Ministère du commerce. Annales du commerce extérieur. Paris. 1843-1851 have title. Documents sur le commerce extérieur. Various changes in name of ministry: 1843-1851, Ministère de l'agriculture et du commerce. 1852-Feb. 1853, Ministère de l'intérieur, de l'agriculture et du commerce. Feb.-May 1853, Ministère de l'intérieur. June 1853-July 1869, Ministère de l'agriculture, du commerce et des travaux publics. Aug. 1869-1881, Ministère de l'agriculture et du commerce. Occasionally published in combined numbers. No more published. Année 1915, fasc. 11-12 forms index to 1843 1915. MnU PER 1840 1842, n. ser. 1909 1917.	2074900	France-Amerique. New York. MnU NEW Current 3 years.	2076300
France. Ministère des Affaires étrangères. Commission de Publication des Documents relatifs aux Origines de la France. Commission de Publication des Documents relatifs aux Origines de la Guerre de 1914. Documents diplomatiques français (1871-1914).	2075000	France États Unis; revue mensuelle du Comité France-Amerique. Paris. The "France America committee." Comité France-Amerique. Paris. Ceased with v. 6, 1924. Merged into France-Amerique Jan. 1925. Contains monthly supplement. France. 20 15182 MnU PER V.1:6 (1919-1924).	2076400
		France illustration. Paris. Title varies: Mar. 1947-1948, France illustration littéraire et theatrale. Absorbed by Nouveau Femina. Supplément theatrale et litteraire. 48 24276 MnU PER OCT1945 1955, suppl. no. 1-191 (1947-1955).	2076500
		France illustration litteraire et theatrale. See France illustration.	2076600
		France Institut national d'études demographiques. Travaux et documents. Paris. MnU XX V.9-11,13-19,21-40,43- (1950- ).	2076700
		France libre; liberte, egalité, fraternité. London, Hamish Hamilton limited. Ceased pub. v. 13 no. 74 (Jan. 1947). MnU PER V.1:13 (NOV1940 JAN1947).	2076800
		La France litteraire. Paris. No more published? 14 21183 MnU PER V.1:36 (1832-1839).	2076900
		Franciscan historical classics. See Academy of American Franciscan historical classics. See Academy of American Franciscan historical classics.	2077000

539

Fig. 1. Minnesota Union List of Serials. Preliminary edition sample page.

provide maximum benefit for Minnesota's library users. Our state's library environment features:

- one large academic research library—the University of Minnesota;
- many smaller academic libraries in the 75,000–250,000 volume class;
- two large public library systems—the Minneapolis Public Library and the St. Paul Public Library;
- one private research library—the James Jerome Hill Reference Library —which serves as a nucleus for the metropolitan area private college

library network called CLIC (Cooperating Libraries in Consortium); and

- some library automation activities among these libraries, with the largest automation staff and activity at the University of Minnesota.

The parallel developments of networking and systems design at the university made possible the proposal to the MINITEX Program Advisory Board for funds to develop the system and publish the first union list. In summer 1971 this program received approval and work was begun in mid-August. On September 1, 1972, the Preliminary Edition of MULS was published and distributed to participating university and MINITEX network members. Following is a report of this work, its results and problems.

### PROGRAM SCOPE

Obviously, to create a system capable of eventually including library holdings state-wide and to convert such data requires definition of an initial and future scope. The initial scope was defined as:

Conversion of the University of Minnesota Libraries' actively received titles, departmental libraries' complete titles, and inactive titles in the Libraries' Periodical Division. Development of a batch input tape software system capable of supporting initial conversion, correction, and updating to produce the Preliminary Edition of MULS.

The future scope would potentially include the augmentation of the MULS data base with the following non-University of Minnesota holdings:

- a. Eight metropolitan area private colleges in the CLIC network, with production of a CLIC union list for their members' use;
- b. Minneapolis Public Library serials and unique titles from other public libraries of over 50,000 volumes, with production of a public libraries union list;
- c. Holdings of all state agencies, which would include the Minnesota Historical Society, State Law Library, State Department of Health, and Legislative Reference Library, with production of a union list for their internal resource sharing;
- d. State supported colleges' holdings;
- e. University of Minnesota inactive general collection serials, thereby completing access to the state's largest research library;
- f. Private college holdings outside of the metropolitan area CLIC institutions; and
- g. Selected special libraries' holdings.

At the moment of this writing we have the initial scope completed, are just completing a, b, and c, and have planned work on d and e for 1973.

In view of this scope the initial MULS magnetic tape system was based on the MARC format to permit:

- publication of a photocomposed or line-printer-method full union list;

- publication of regional combination or individual library lists using an IBM 1403 line printer equipped with the ALA graphic print train;
- storage of complete and verified information on each serial as known, together with the source of the cataloging data;
- extraction of the data via individual libraries to assist those wishing to develop automated serials management systems including check-in, claiming, binding, etc.;
- conversion of the file to other storage media such as disk;
- fulfillment of the smallest to the largest libraries' needs for bibliographic detail; and
- extension to a fully automated resource sharing system which would further improve the benefits of library cooperation.

With this picture of the program scope, the design factors, data conversion, computer system, programs, photocomposition, costs, and problems will be described below.

## SYSTEM DESIGN

The easiest way to look at the MULS design is to gain an understanding of the MULS MARC Record content as shown in Table 1. This record is the basic unit which is entered, including all associated cross-references or added entries to be made. It in turn generates each of these secondary entries in the file. In this brief description we will assume the reader is familiar with the MARC serials record as described in *Serials: A MARC Format: Preliminary Edition* and its *Addendum No. 1*.<sup>1, 2</sup> There are some differences between the MULS format and the LC MARC format, most importantly the addition of a sort field (Tag 249) and the subfield arrangement for holding fields (Tag 850). Other variations have been indicated in Table 1, which uses the same organization as that contained in the LC format description referred to above.

Figure 2 shows a page from a master-file listing. Note entry no. 2074000. This listing is formatted with the sequence number of the record appearing on the first line, followed by the bibliographic level and the remaining leader information. Next the record directory entries are found for fields 008-950 as applicable. On the next line are the 008 fixed length data ele-

Table 1. MULS MARC Record Content

- A. Leader
1. Logical record length—five characters
  2. Record status = 1 for MARC record  
= 4 for added entry (AET)  
or cross-reference (XRF)  
entry
  3. Legend
    - a. Type of record—not used (blank)
    - b. Bibliographic level = s
    - c. Two blank characters
  4. Indicator count = 2

5. Subfield code count = 2
6. Base address of data = 5 characters
7. Sequence number = 7 characters
- B. Record directory
  1. Variable field tag = 3 characters
  2. Field length = 4 characters
  3. Starting character position = 5 characters
- C. Control fields—008 Fixed Length Data Elements
  1. Date typed
  2. Publication status
  5. Country of publication code
  9. Type of serial designator
  10. Physical medium designator
  12. Form of content
    - a. Type of material code
    - b. Nature of contents codes
  13. Government publication indicator
  14. Conference publication designator
  20. Language code
  21. Modified record designator
  22. Cataloging source code
- D. Variable fields
  1. Indicators

In general we have not followed LC in the use of indicators. One exception is the use of filing indicator for the 100 and 200 series tags, which we implemented before seeing that this feature was provided in the Addendum No. 1 to the LC format. Therefore, the indicators except as above are both blank.

2. Subfield codes

Except for the holdings statements (TAG 850) we have generally followed LC philosophy. For TAG 850 we now precede the \$a subfield with a \$z subfield, suppressed on printing, which contains the 4 digit number identifying each specific holding library which is also found at the end of the 008 field.

3. Variable fields currently used.

- 010 LC card number
- 022 Standard serial number
- 041 Languages
- 100 Main entry—personal name
- 110 Main entry—corporate name
- 111 Main entry—Conference or meeting
- 200 Title as it appears on piece
- 245 Full title
- 249 Sort key from 100 or 200 series tags stored in collating codes and limited to 120 characters
- 250 Edition statement
- 260 Imprint
- 500 General note
- 501 Bound with note
- 515 Note for explanation of dates, volumes, etc.
- 525 Supplement note
- 555 Cumulative index note
- 730 Added entry
- 850 Holdings
- 950 Cross-reference tracing

NOTE: We have followed LC numbering for the above data elements, and have substituted blanks on the tape record for those elements omitted. We have also expanded the 008 field to include a variable number of 4 character elements which contain the index number of each holdings location listed in the z subfield of TAG 850.

```

2073990<022200133003621 00R00400000 11000510004D 245004000100 249008200140 260001100222 500005300233 730004600286 850005800332
ASGCGJ80(C3H)
71122a fr c y frz m 0133 0130
$a France. Ministère de l'éducation nationale.†
$aa Annuaire de l'éducation nationale.†
$ab FRANCE--MINISTERE-DE-LEDCATION-NATIONALE--ANNUAIRE-DE-LEDCATION-NATIONALE.E
$ab Paris.†
1$a "Publié par l'Institut p'edagogique nationale."†
1$b France. Institut p'edagogique nationale.†
1$c MARC Subfield 1962,1965,1968,1970- .#eJ70.58 qP844†
2$a MARC 258 $cdCurrent volume only.
2074000<022200217016431 00R004500000 010001800045 110005000063 245003800113 249008200151 260001100233 500044000244 500005400684
50001720073H 50001470051U S15007901057 730006901136 730005301205 730005401258 730003801312 850007601350
71122a fr c I frz c 0107
$aa 0-445IN rev 2†
$aa France. Ministère de l'éducation nationale.†
$aa Catalogue des theses de doctorat.†
$ab FRANCE--MINISTERE-DE-LEDCATION-NATIONALE--CATALOGUE-DES-THESES-DE-DOCTORAT.E
$ab Paris.†
1$a At head of title: 1884/1885-1931: Ministère de l'instruction publique; 1932-1958 minist'ere de l'educatio n nationale;
1960-1968. 1884/89-1958: Issued annually, 5 yearly parts being pagged continuously and supplie d with a general to p- and
subject index to form a volume. Vols. 1-5, 8 (1884/89-1904/09, 1919/23) have also an alphabetical list of authors. In
vols. 6-7 (1909/13-1914/18) a list of authors is given with each fascicle.†
2$a Beginning with 1959, vol. numbers no longer used.†
3$a Beginning with theses of 1930, a list was also published in parts, as supplements to the numbers of the Bibliographie de
la France, 1. ptic., Bibliographie, 1932-date.†
4$a Title varies: 1884/89-1958, Catalogue des th'esen et 'ecrits academiques. 1965- ; called Bibliographie de la France;
Supplément Annuel D.†
1$a 1915-16 (fasc. 32-33) and 1917-18 (fasc. 34-35) each issued in one number.†
1$b France. Ministère de l'instruction publique et des beaux-arts.†
2$a France. Direction des biblioth'eques de France.†
3$a Bibliographie de la France. Supplément Annuel D.†
4$a Catalogue des theses de doctorat.†
1$a MARC Subfield V.1-15 (1884/1889-1958); n. ser. 1959- .#e016.J78 qP84
2074100< 2200061002824 00R000500000 150014800005 149006700154
0107
$a France. Ministère de l'instruction publique et des beaux-arts.$bFrance. Ministère de l'éducation nationale.Catalogue
des theses de doctorat.†
$a FRANCE--MINISTERE-DE-LINSTRUCTION-PUBLIQUE-ET-DES-BAUX-ARTS.#J
2074200< 2200061002474 00R000500000 150014400005 249006700119
0112
$a France. Ministère de l'instruction publique et des beaux arts.$bBulletin des bibliothèques et des archives.†
$a FRANCE--MINISTERE-DE-LINSTRUCTION-PUBLIQUE-ET-DES-BAUX-ARTS.#J
2074300< 2200061003914 00R000500000 150021000005 249011500215
0129
$a France. Ministère de l'instruction publique et des beaux arts. Comité des travaux historiques et scientifiques.$bFranc
e. Comité des travaux historiques et scientifiques. Section de Géographie.Bulletin.†
$a FRANCE--MINISTERE-DE-LINSTRUCTION-PUBLIQUE-ET-DES-BAUX-ARTS--COMITE-DES-TRAVAUX-HISTORIQUES-ET-SCIENTIFIQUES.#J

```

Fig. 2. Master-File Listing.

ments with the last four digits the holdings location index number which is the same as the suppressed \$z subfield in the 850 field. Then the variable fields are listed in numeric sequence. Note the subfields as indicated by \$z, \$b, etc. The number to the left of each \$a is the MARC tag number. Another departure from MARC is to store the call number as a subfield of the holdings statement since it may vary among participating libraries.

To contrast how the information is stored and how it appears when published, the same record is shown in the left column of Figure 1. Also, the next record shown is generated from an added entry TAG 730 in this parent record. We have prepared a detailed coding manual which is followed by our coders; this document presents various examples of conditions and details the full system structural requirements.

These changes in the format were made to simplify wherever possible, to provide for conditions which the original LC format did not cover, and to preserve the MARC structure with full text. With the exception of subject headings, all bibliographic text is stored. Other MARC tags may be added to the system at any time.

The initial system was tape-based, as our computer system at that time did not have uncommitted disk drives. Also, we needed to gain some detailed knowledge of the file and record characteristics to most effectively design the disk-based system. This knowledge could be gained easily after

some basic data were stored in the system. Since programmer time was our most precious commodity, this phased approach was used to: (1) achieve enough support on the tape system to permit publication of the Preliminary Edition of MULS while gathering file and data characteristics; and (2) bring into operation a disk-based system with completely automatic added-entry correction and generation, coupled with very flexible correction procedures.

## DATA CONVERSION

Various methods of data conversion were investigated. Two requirements seemed obvious in our system—compilation of data on a code sheet and efficient, accurate keyboarding. Further, since the MARC character set was being used, any potential device had to provide a minimal keying situation to accommodate this character set. Compilation of data on a code sheet was necessary because multiple files in multiple locations would be checked to gather all of the information. Keyboarding had to be efficient as it was initially estimated that some 25 million characters would be entered before we were ready to publish the union list.

The IBM Model V Record Only Magnetic Tape Selectric typewriter (MT/ST) was chosen as offering the best approach for high volume, short duration use. Three machines, each equipped with the special MARC element and key buttons, were leased. Typists easily corrected their discovered errors on these units. Each typist followed detailed typing instructions and, after mastering the coding manual practices and procedures, was a trained coder.

During July/August 1971 all training aids were prepared, forms designed, and staff recruited. The initial staff complement received their training during the last two weeks of August. During September the data gathering staff was brought to full strength and consisted of:

Project Director	1 FTE
Editors (librarians—library assistants)	4 FTE
Senior Clerk-Typists	6 FTE
Clerks (students)	12 FTE

Full-time equivalents are used as staff were in many cases part time or temporarily lent to the project. During the period August 1971–June 15, 1972, which comprised the total data preparation time for the Preliminary Edition, five librarians and thirty-five students actually were trained and participated in the project.

It took about six weeks to bring most of the staff to an acceptable performance level. Some students found the work too complex or detailed and voluntarily left the project. One clerk-typist did not gain sufficient proficiency to pass out of a trainee status and was terminated at the end of her probation period. Thereafter, with a staff of this size, performance problems were minimal.

The data to be included in the Preliminary Edition comprised the university's

- currently received, centrally recorded serials (20,000 titles);
- inactive Periodical Division titles (8,000 titles);
- coordinate campus locations of the university (4,000 titles);
- complete departmental library titles excluding the Bio-Medical Library (6,000 titles).

The Bio-Medical Library was excluded due to its present mechanized serials system which would be used to produce a separate serials list, issued as volume 3 of MULS to the university and the MINITEX participating libraries. This separate publication was necessary due to the short time in which the initial data were to be collected. However, the Bio-Medical Library is now also being included in the body of the MULS data base.

These four categories of serials necessitated quite different approaches dependent upon the available check-in files, shelflists, or catalogs. For example: to capture data on the currently received, centrally recorded titles we photocopied the Kardex drawers from the serial check-in file maintained in our headquarters library. These running titles were checked against the official card catalog in the library. If the title was found, the bibliographic information was transcribed, together with all Kardex and catalog locations. If not, the Kardex data were copied onto a code sheet for subsequent verification together with its listed location. About 5 percent of the time the photocopied sheet was illegible. These entries had to be transcribed from the check-in file, verified, and then passed on to the next step. When bibliographic data had been assembled on the code sheets they were edited in groups, each group accompanied by its photocopied sheet. Corrections were entered by editors, the catalog or check-in file was rechecked as necessary, and then the sheets were sorted by holding location.

Next all holdings information was procured from the remote location to make sure it was the most reliable information. Finally, the sheets were returned to be rechecked and typed. "Mopping up" occurred at each holding location to encode inactive titles and uncataloged serials. When a title could not be verified, the piece itself was used to develop the main entry, added entries, and other pertinent cataloging information.

Similar procedures were used on the inactive Periodical Division shelflist. Departmental library locations involved the use of shelf-locator visible indexes and shelflists, coupled with check-in files and branch catalogs. Coordinate campus locations outside the Twin Cities metropolitan area required the checking of title/holdings listings provided by these campus libraries. Many entry problems resulted, because variant cataloging approaches were used in many of these libraries.

Typing and subsequent input were done as coding sheets became ready for keyboarding and were therefore in random order. Over 40,000 individual records were typed, each averaging about 480 characters (an approximate

18 million keystrokes). During the period February-June 15, 1972, when the complete file was proofread from the thirteen volume master-file listing, another 5 million keystrokes were required to delete, to reenter, and to correct entries and associated cross-references. Our final keyboarding stroke count was exceedingly close to our original estimate of 25 million characters.

The proofreading portion of the data conversion took twice as long as originally anticipated, causing a delay of two months in photocomposition scheduling. Proofreading was completed on June 15, 1972, and on the following Monday the photocomposition vendor received the final output tape. Due to some format changes and continued systems problems the photocomposition output was not received until July 21. Printing and binding followed and on August 28 the Preliminary Edition, consisting of 1,566 text pages in two class A bound volumes, was ready for distribution.

### COMPUTER SYSTEM

Two computer systems were used in MULS production. One system was used to convert MT/ST cassette tapes and involved initially an IBM 2495 cassette converter coupled to an IBM 360/20 system. This configuration was replaced by off-line tape conversion using a Data Action Tape Pooler and the same computer for code conversion and record blocking. Two-hour to one-day service was provided by this service center, located in a local insurance company.

The raw data tape resulting from the above process then required processing on the second computer system, an IBM 360/50 at the University of Minnesota. All programs are written for the COBOL F compiler and operate under OS/MFT using 1600 cpi magnetic tape. Two 80K core partitions are required for the updating and printing programs. The ALA graphic print train is used to print the file and control listings. Figure 2 was printed with this character set.

### PROGRAMS

MULS programs for the present tape system were conceived as two sets: (1) conversion, file creation, and updating; and (2) printing functions. The first set performs the following functions:

- identification and checking of fields for validity, tagging, and structure from the raw input tape;
- creation of MARC-type main entries;
- creation of secondary entries generated from the added entry (TAG 730) and cross-reference (TAG 950) fields;
- creation of correction and deletion entries;
- sorting of main entries and the generated secondary entries in alphabetical sequence;
- sorting of correction and deletion entries in sequence number order;
- addition of new records;

- deletion of an old record;
- addition of a new variable field, including holdings statements;
- substitution of data in a variable field;
- deletion of a variable field;
- production of a transaction file reflecting changes to the data base; and
- generation of a new master tape, which can include resequencing the entire file and/or producing a work list of the file.

However, any change in a 100, 200, 730, or 950 tag requires deletion of the complete record with its secondary entries, and reentry of the record in its changed form. This is because a two-pass update would be required in the tape system to automatically correct secondary entries as well as to generate them.

The second set of programs performs the following functions:

- printing of a formatted work list selectively by location or combination of locations, diacritical printing preceding the character to which it applies; and
- printing of a conventional union list format which closely duplicates the design of the photocomposed page in Figure 1. Selectivity by location or groups of locations is present and all diacritical characters are overprinted as in the photocomposed list.

## PHOTOCOMPOSITION

The Preliminary Edition of MULS, as shown in Figure 1, was photocomposed by a Twin Cities firm using a Harris Fototronic CRT composition system and an IBM 370/145 computer system. We chose the lowest bidder which was fortunately a local firm. The bid required the vendor to program from our MARC format master file tape an input tape for the photocomposer which would produce the specified format, using the MARC character set in a font to be chosen from sample text pages. The vendor's bid included programming, composing, and procurement of several of the characters used by MARC which were not in his current font repertoire. A test tape was provided to the vendor for his developmental use, together with documentation on the MARC MULS system.

After seeing the initial result of our specified format we were not pleased with the result. The reason for this was compounded by the fact that:

- the vendor had not followed some of the suggestions;
- the vendor had made some unspecified changes;
- the program had injected some data errors and other unacceptable conditions; and
- the library, in its total lack of experience with this variable density form of display, had no idea of the real effect of its proposed format in getting efficient character density coupled with attractiveness.

Each of the design problems was looked at in order to adjust character

size, column length or width, continuation line placement, display form (bold, regular, oblique), and relative data element placement. Four iterations were required to finally produce the format shown in Figure 1. As a result, our photocomposition and printing costs were half the costs had the original format been developed. Style and readability also improved dramatically.

The choice of type font was made by comparing sample pages in both serif and sans serif styles, including Times Roman and other well-known fonts. Various library staff members were asked to vote on their preferred font. News Gothic was an overwhelming favorite by both public and technical services oriented librarians.

The photocomposition vendor had produced many catalogs and books using other special alphabets and characters, but had not previously done any catalog from a MARC format tape. This made possible a high degree of expertise on their part in handling our special character requirements, but added some developmental problems because of lack of MARC format experience. Except for superscripts, subscripts, and the underline, all MARC characters have been needed to display the text.

Our advice to those considering catalog photocomposition is to request bids, as the price on this service has continued to drop. The page price will be dependent upon the services performed. In our case the vendor handled all composition programming. One can estimate that at a minimum 40 percent-50 percent of the page charge would be involved in this service. Also, the size of the job will cause a variance in the price a vendor will quote—the larger the number of pages, the cheaper the cost per page. On a very large application it may be to the library's advantage, if resources permit, to train their own programmer to program the composing device. However, we feel that our best needs were served by contracting for this support as our programming staff was limited and did not have any prior composing-machine experience.

## COSTS

The expenditure to produce a computer-based serials catalog will vary dependent upon salary and equipment rates and the conditions found in the library system. In the case of MULS, condition of the files used ranged from disastrous to excellent, yet with only fragmentary information in each file. Moreover, entry forms varied greatly among the many check-in, shelf-list, and catalog files. Therefore, data collection was much more expensive than it would have been had we keyboarded directly from one existing file of data. To present some idea of costs for others planning similar activities, we have developed some average costing information from our expenditures.

Each main entry in MULS costs \$2.81 on an average, figuring all known actual charges or subsidized costs. This main entry cost includes all associated secondary entries, which is about one secondary entry generated per

1.5 main entries. This \$2.81 breaks down to approximately \$1.00 for design, programming, and administrative costs; \$1.40 for data conversion; and \$.41 for photocomposition, final printing and binding.

Let us look at some specific items which figure into this average cost per record to give the reader some idea of what is reasonable to expect in a project of this sort. A good example is conversion of MT/ST cassette tapes to computer compatible magnetic tape, including code conversion and blocking of the records. Our per-cassette conversion cost varied from \$.50 to \$2.00 per cassette. This variance was caused by a change from on-line to off-line conversion and the problem of handling cassette tapes which did not have the proper stop code at their end. Our actual billed average throughout the whole project was \$.73 per cassette. If no tapes had been prepared omitting stop codes and if total off-line conversion had been used, our average would have been \$.50 per cassette. A typical cassette tape averaged seventy-five new MARC entries, so this was a very economical charge for this method.

Another specific cost to examine is computer time. On our IBM 360/50 system, time is billed as time on/off the system and not according to some calculation of CPU/channel/storage/peripheral device usage. Normally an internal university rate is a great deal cheaper than a commercial rate for the same equipment. However, the billing method used in our system has probably increased our costs for computer time over the CPU time method of billing, since the user is at the mercy of contending with other jobs on the system at the same time; i.e., waiting for his processing turn. This has had a noticeable effect in our case; run times to update the file have varied from four to six hours machine on/off time almost independent of the number of transactions being processed.

Photocomposition page rates over the last few years have been dropping as competition in this area has flourished. Two years ago it was common to receive quotes of \$6.00 per page or even higher. Most prices we received were under this figure; but at the time our contract was signed, our successful bidder, who also was our lowest bidder, quoted \$2.60 per page. This included full programming support to convert our MARC format tape for creation of the photocomposer input tape. Today rates much lower than this can be found. Moreover, rates under \$1.00 per page can be obtained if the customer is able to create his own input or driver tape for the photocomposition device, making this method considerably more attractive for even low volume per-page printing. In the case of MULS, one photocomposed page equals ten double column computer printed pages without photoreduction. Photoreduction can cut computer output pages about one-third, yet obviously not to the limit achieved through the photocomposition method. Therefore, considerable printing costs can be saved dependent upon the number of copies of each page printed.

## PROBLEMS

The problems encountered during this project and its daily operation presently have been, for the most part, those commonly found in any large scale project. The large volume of data, less than ideal computer environment, condition of the original data, and large staff required to produce this effort all magnify many problems which seem unimportant in a small or short term project. In general these problems fall into the following categories: (1) data handling and bibliographic; (2) communications; (3) estimating; and (4) hardware or computer related problems.

*Data Handling and Bibliographic.* Those who create and use research library catalogs can appreciate the formidable physical problem in any data conversion activity. A half century or more of cataloging variations must be brought together; mistakes in the original data, differences in format of cards, and spelling or usage inconsistencies must be weeded. Couple this situation with a new staff, large in number but containing few professionals. The result could be disastrous if proper decision-making and problem identification did not occur. Not knowing the magnitude of these problems we decided on almost verbatim transcription of records but spelling out all abbreviated words in any filing field.

When our first file listing appeared—some 40,000 main entries plus 30,000 secondary entries—we saw that the filing arrangement was very poor due mainly to spelling variants, failure to consistently follow instructions to spell out abbreviated terms (which somehow escaped editing), and different entry forms for the same body. Transcription of data from the original source was very accurate but because of these problems in the original data our proofreading resulted in some change occurring in about 10,000 of these 70,000 records. The use of punctuation marks in main entries varied so much that some corporate entries were filing in five or six separate groups in the list, each separated perhaps by several pages.

The great shocker was the arrangement under the United States, as some coders had copied exactly from the card without spelling out U.S. and inserting a period and space. About a dozen entries had failed to be caught by the editors and appeared as one block. Then, to compound the problem, others spelled out United States but forgot to insert a period after it. Moreover, very early in the project the typists incorrectly inserted 2 spaces after the period. In all, there were six forms to the U.S. entries alone, with only one being correct.

This lesson taught us that no matter how well instructions and examples are prepared misunderstanding can result; and, of course, editors and others will not catch all possible errors. However, these major errors were eliminated before publication. With the large volume of data and limited funds our conversion process was quite streamlined with most of the error-checking resulting after the data were on tape and displayed in their proper relation to other records. Few keyboarding errors occurred which

were not caught at typing. The predominant errors resided in the nature of the original data, or in the lack of some piece of information from three or four different files which may have been checked in building the full record.

*Communications.* In any large project effective communication is necessary to improve quality of work and progress toward completion of the scheduled task. Frequently scheduled meetings of the staff were used to inform all project members of decisions, receive their suggestions and criticisms, and develop coordinated work assignments among the teams of each editor/librarian. All typing personnel were trained as coders and were periodically relieved of typing to code. This gave them an insight into detecting problems for referral to the professional staff, renewed their knowledge of proper format, and provided more variety in their work. All project members were capable of performing tasks of coding, control list checking, and proofreading. The most capable clerical staff also assisted the editors in editorial work. It was felt that our use of the team approach, unified training, frequent staff meetings, and very detailed written documentation served to channel communication with a resultant minimization of these problems—once the first few months of the project had passed.

*Estimating.* In most data conversion work accurate estimating is required on many matters. Some estimates we made were very accurate, such as basic time and staff to complete initial coding, typing time and staff, and supplies needed. However, other estimates were not very accurate. For example, the time to edit and correct the file once basic data collection was completed was double our original estimate and required more typing than anticipated. This caused the publication schedule to be delayed two months. Difficulties at the computer center and at the photocomposition vendor caused another two months delay, even though it is doubtful that our photocomposition firm would have been ready had we met our original estimate. Our original target was publication not later than two months after the basic data collection period of six months, i.e., in eight months. However, on a project of this size, and with the addition of about 7,000 more titles than we had originally estimated, we did not feel that the fifty-four weeks really required was excessive.

Computer time was also difficult to estimate because of the time on/off the system. Dependent upon the nature of the other jobs on the computer, this time varied greatly, for updating runs were almost independent of the number of transactions. There is always room for improvement in estimating, and, obviously, we have learned many things from this experience to use in further work.

*Hardware/Computer Center.* Our largest problem was creating firm computer scheduling commitments on our campus IBM 360/50 computer, which serves the business functions of the university. All other campus computing facilities use Control Data equipment which is six-bit character, word oriented. With the extended character set requirement and the availability

of the IBM 360/50, which we were already using for other work with the ALA graphic print train, it was natural for us to choose this system. Current facilities are now satisfactory to permit our tape batch system operation and the development of our new disk-based batch system. Tape pooling operations for the MT/ST have caused some problems due to equipment changes at our vendor. We have now switched to a new conversion source as our former vendor upgraded his data entry system to key-to-disk. The three MT/ST typewriters we leased performed quite reliably, but one machine seemed to have more down time than the others. Now that our typing load is down, we have cancelled two Model V's and will maintain two machines. We are now choosing a new system for key input to cassette tape. On the new equipment we will do our proofreading and initial correction off-line resulting in a further cost saving. This was not possible previously as our typing load required two-shift operation on all machines during the Preliminary Edition preparation time.

## CONCLUSION

A great amount of effort has been expended to achieve a unified serials data base to serve Minnesota's libraries. It is our hope that this system can continue to be developed in as flexible a way as possible so that future needs can be supported through the system. Only the imagination of those involved in networking is the limit to identifying the future needs to be met through access to this data base. Of course, we would hope that one day our data could benefit the development of other similar programs in other states and, perhaps more importantly, in achieving a true national serials data base.

## ACKNOWLEDGMENT

Many staff members at the university and other institutions contributed their invaluable counsel as we have proceeded on the development of the system and the data base. The MULS project staff particularly receives our deep gratitude for its yeoman effort. Special commendation is due Mr. Don Norris for systems design and principal programming support. Mr. Carl Sandberg, who wrote all printer output programs, also contributed invaluable assistance to the project. The MINITEX program and University Library administration receive our appreciation for placing their confidence in the Systems Division. MULS and its support system is truly a product resulting from the coordinated concern and interest of the aforementioned individuals and groups.

## REFERENCES

1. U.S. Library of Congress. Information Systems Office. *Serials: A MARC Format*. Preliminary edition. Washington, D.C.: Library of Congress, 1970 (L.C. 73-606842).
2. U.S. Library of Congress. MARC Development Office. *Serials: A MARC Format. Addendum No. 1*. Washington, D.C.: Library of Congress, June, 1971.