LC/MARC CD-ROMからのデータ変換

Data Conversion from LC MARC CD-ROM

小 松 泰 信*

羽衣学園短期大学図書館

Komatsu Y. : Data Conversion from LC MARC CD-ROM. *Igaku Toshokan* 1989 ; 36 (2) : 89-102.

Data downloaded from a database is assumed to be transferred to another system. This is effectively performed only with appropriate methods of data processing at downloading, data conversion, generalization and utilization. An actual example is presented in this paper. The author conducted the transfer of data from the LC MARC CD-ROM database to a company database run on a personal computer and to a mainframe, the HITAC/L-470X. The process is detailed in this paper using the standardized BASIC interpreter. The author also depicts the real course of batch loading onto a university library system, as well as discussing the independence of bibliographical data from a library system.

I. はじめに

データベースをCD-ROM (Compact Disc-Read Only Memory) 化したものの中で、ダウンロードの機能を持つものは、データ変換による再利用あるいは、何等かのシステムへのアップロードを想定したものであるといえる。ここに視点をずえるなら、他システムとの何等かのデータ上の互換を計るのは、必然の成り行きであろう。

本学では、某大学社会学部より7,500冊余りの洋書の寄贈を受けていた。この遡及入力作業を対象に、LC/MARC CD-ROM である BiblioFile からの書誌データダウンロードとオリジナル入力、およびそれら MARC レコードのメインフレームデータベースへのデータ変換を試みることにした。

CD-ROM というメディアの出現で、手軽に MARC と接することが出来るようになった。データ変換とその利用も、より平易で実用に値するものを、自ら開発することがCD-ROM の今後の利用形態にも好ま

しい影響を与えよう。

本稿では、①CD-ROM を用いたデータ作成の過程と②データ変換の実際について具体的に述べ、さらに③CD-ROM のシステム上の可能性および④ソースデータとローカルシステムの独立について考察してみる。

II. CD-ROM メディアからの変換

1. その意義と利用の現状

書誌情報データベース (以下DB) 型のCD-ROM は、それなりの普及を示し始めたようである。しかしその利用の実情は、パーソナルコンピュータを用いたシステム下でのスタンドアロン (オンライン化あるいはネットワーク化されていない単体) によるものが多い。既にトータルシステムを持っているところでも、システムの流れから外れたところで発注・選書や目録図書カード作成用に、或いは利用者の検索用としての利用にとどまっているようだ。

多くの海外DBが、あいついでCD-ROM 版を発表するのもオンラインの補助的機能としてCD-ROM を位置づけているためであろう。

そうした媒体のひとつである BiblioFile は、LC

* Yasunobu KOMATSU : Hagoromo-Gakuen Junior College Library, 1-89-1 Hamaderaminami-cho, Sakai-City, Osaka 592, Japan.

(平成元年1月31日 受理)

／MARCのうち、Complete Service・GPO・Canadian（単行本・雑誌・政府刊行物・地図・視聴覚資料・カナダ出版物）の約370万件（1988年現在）を、CD-ROM化したものである。データ内容は、1965年から現在までに納本された網羅の文献情報すべてと出版予定のCIPデータ及び、週及入力による1900年以降の一般的文献書誌情報が含まれる。

CD-ROMから得られたデータを加工して利用したり、ユーザーのデータベースを構築することは、誰しもが考えることである。しかし、データ作成という観点に立てば、1件ずつの書誌データに煩雑なローカルデータを追加して、やっとダウンロードするのでは、入力ツールとして見ても効率的とは言い難いであろう。今までのような、磁気テープ（以下MT）を利用し、大型機での何千件という突き合わせ作業とデータ追加処理を知るものにとっては、その非効率性を指摘するのはやさしい。しかし、そこから得られる書誌データの質は、直接入力と比較してはるかに均質なものである。個々の図書館がCD-ROM版DBの利用技術を確認する意義は大きいだろう。

ところで、CD-ROMからいかに大量のダウンロードが可能であっても、自館DBに落とし込むには、何処かで対応した所蔵事項等のローカルデータの入力作業をおこない、それを突き合わせ結合させなければならない。

CD-ROMが出現するまでは、書誌データとローカルデータにそれぞれ、図書番号・Bar-Code No.等のAccession No.をユニークキーとして磁気ディスク上での突き合わせ作業が、多くの場合大型計算機を用いて行われてきた。結合の出来たデータは、MTで供給される。大型計算機を利用しMTを使うこの作業に比べ、CD-ROMは、見劣りするものであろうか。

拙稿は、スタンドアロン状態にあるCD-ROMドライブ接続のIBM PC／AT Compatibleパソコン東芝J-3100から、バッチ（一括）処理によりフロッピーディスクを介して、本学のホスト図書システムであるHITAC L-470Xとスタンドアロンの状態にある各種パソコンDBへのデータ変換の実際とその過程を詳説するものである。その過程で、データの流れを

念頭において異なるハード構成の場合、CD-ROMがどういった役割と機能を担うことが出来るか考察していきたい。

2. 効率的ダウンロード

一口にダウンロード機能を利用するといっても、次のような問題を考慮しなければならなかった。

- ①ダウンロードしたデータと付加したいローカルデータの結合
- ②ダウンロードしたデータのファイル形式
- ③コンバート用のファイル出力機能

ここで、①はCD-ROMから抽出してきた書誌データに、こちらの付加したい所蔵事項等のローカルデータをいかに効率よく結合して行くかである。

②③に関していえば、以下のような点が問題となる。ダウンロード出来るDBの大半は、データ＝コンバート用のデータで、テキスト形式のASCIIセーブされたデータである。ところがこのデータを検索して再び編集することは、出来ない場合が多い。さらにデータの再利用を考える場合には、変換用のプログラムを組まねばならない。ただ、データの形式としてテキスト形式¹⁾でセーブされたものは、パーソナルコンピュータで扱うには比較的容易でもあるため、標準的形式としてそれに慣れておけば応用がきくであろう。

一方、メインフレームを持つところでは簡便な方法として、データを大型機で直接加工し利用する方法がある。LC／MARCというCommunication formatに該当する可変長形式のままで可能である。一括・大量にCD-ROMからデータを吸い上げ、加工したものをメインフレームDBに落とし、同時にパソコン等に分散処理用のデータとして出力する方法も考えられるが、この方法は、後に述べるように、むしろ現状の一般的なCD-ROM利用規定に抵触しないか、十分検討される必要があろう。

筆者は、CD-ROM利用の現状を考慮し、メインフレームのみでのデータ加工を避け、パーソナルコンピュータ（以下PC）上で変換を試みることにした。加えて、PC上で使用可能なデータが、メインフレームにも乗るものにして、二重に利用可能な方式を取った。

図1. バッチ処理フロー

図2. バッチ変換処理におけるシステム構成

この方法だとメインフレームを持たないところでもデータ変換が可能となる。また、メインフレームを持ちながら、P.C.でのデータマネジメントが遊離化している場合の、データの有機的活用を試みてみた(図1)(図2)。

3. 対象となる文献の傾向とダウンロードの実際

今回対象となった図書館の特色に言及しておきたい。

- 分野としては、社会学を中心に社会科学系の文献が約68%を、人文科学系が30%をしめている。
- 次に、複本が多い。テキストとして利用されたので

あろうか、40～50冊にもおよぶ複本が少なくない。

逆に、全集・シリーズ・多巻物はわずかであった。

- 出版地はアメリカが80%強、イギリスが15～16%、カナダ出版物も散見できるが、独仏その他の出版物は殆ど見受けられない。

- 最後に特筆すべき点は、出版年が1930～1950年代のものが多くことである。しかも reprint や版次の違うものが散見される。この年代は、LC/MARC が十分に整備されていない年代で、今回の作業の性質を特徴づけた。このためデータの利用は、ヒット→ダウンロードだけでなく、LC準拠のオリジナ

ル入力と別版次でヒットした MARC の書誌修正
→ダウンロード出来る機能を必要とした。

3-1. 検索とローカルデータ結合

BiblioFile の場合、追加したいデータの空のタグを用意出来るだけでなく、事前に設定してあるローカルデータについては自動付与できるため、資料種類別に分けて自動付与していった。作業行程としては、CD-ROM から検索→ローカルデータの自動付与→ユニークキー等を付加して完成したものをダウンロードするという流れになる。

このローカルデータのレパートリーをフロッピー単位で幾種類か用意すれば、検索にかける資料あるいはリストをローカルデータ別に並べ変えることで、入力の必要なローカルデータも、登録番号等の Accession No・請求記号・登録年月日といった最小限の入力で済ませることが可能となる。なお、フロッピーの入れ替えなしにローカルデータのレパートリーを99種類まで設定して選べるシステムも提供されている²⁾。

3-2. オリジナル入力

一方、ヒットしなかったものについては、ローカルデータと呼び込んできた入力画面(図3)に、書誌と前記のユニークキーを含む可変項目を入力して、オリ

Int lvl:	Repr:	Cont:	Govt pub:	Conf pub:	Festschr:
Indx:	M/E:	Fict:	810:	Lang:	eng Mod rec: Source:
LC Card	010	ah188			
ME:Pers Name	100	10	a		
Title	245	1	a		
Imprnt	260	abc			
Phys Descript	300	ac			
Note:General	500	a			
Note:Content	505	a			
Subj:Pers	600	a			
Subj:Topical	650	a			
tourokuban	952	a0			
tourokubi	953	a610618			
gousyacode	954	a7320			
KAKAKU	955	a0000000			
ukeireKAKAKU	956	a0000000			
NDC	957	a			
KAN CODE	958	a01			
LOCAL DATA	962	a02b03c01d30e01f61g8000h84			
WAYOUKUBUN	976	a02			
sitanakubun	977	a02			
MARCkubun	978	aLC			

ローカルデータ
付与項目

F5 Original Input End of Record line 23 of 23

図3. オリジナル入力画面

ジナル入力を完成した。このオリジナル入力は、入力画面を自館で独自に設定出来るため、本体への直接入力と比較して、3〜2.5倍の入力効率をあげることができた。オリジナル入力のデータ形式は、CD-ROM からヒットしてダウンロードしたデータと同様の、MARC format になる。

3-3. ダウンロード後のレコード操作

ダウンロードした時点でのレコードは、検索 Key を持っていて、再度検索・編集の可能なファイル形式である。書誌データ全体がフルスクリーンエディター上に展開されるため、再校正やレコードの流用も可能で、例えば版次等が違うだけの書誌については、書誌内容を一部修正して流用した。

完成したレコードを Communication MARC format にしてバッチ出力する。出力された file はデータの append (追加) がきくため、継ぎ足していけばオリジナル入力を含めた数千の LC Communication MARC レコードが出来る。それを保存して、ローカル DB から独立した自館用データを作成した。

Communication MARC format は、前記の LC No を Key に持ったファイルに再構成出来る³⁾だけでなく、Title, Author, LCCN, ISBN, ISSN, Control No の検索 Key を持ったファイルとして、複本調査や将来の自館所蔵データの CD-ROM 化に役立てることができる³⁾。また、Catalog Maintenance System のユーティリティを利用して、ファイルからローカルデータを一括して分離したり、特定のローカルデータのタグにインディケータを付与したものを Future Tag として残し、それ以外のローカルデータを一括して分離することも可能である。

この書誌データと所蔵データの分離機能によって、オリジナルデータの共同分散入力も実現可能となる。

III. バッチ処理によるデータ変換について

1-1. 全体的変換の手順

BiblioFile は、MS-DOS 上で変換用にアスキーセーブされた可変長 LC/MARC Communication

```

000330001983000028002314900017002595040034002765000020003105200128003306500033004
58650001200491952001100503953001100514954000900525955001200534956001200546957000
600558958000700564962003700571976000700608977000700615978000700622      87013805 /
AC 870505s1987      nyua      j b      00110 eng      a      87013805 /AC      a0531103781 0
aTK7882.C56 bL36 1987 0 a004.5/3 219 10 aLampton, Christopher. 10 aCD ROMs /
cby Christopher Lampton. 0 aNew York : bF. Watts, c1987.      a96 p. : bill. ; c2
3 cm. 0 aA First book a"A CD-ROM library": p. 87-90.      aIncludes index.
aExplains how CD ROM and optical information storage work and discusses their ap
plications and possible future developments. 0 aCD-ROM xJuvenile literature. 1
aCD-ROM.      a002020      a610618      aZ320      a0000000      a0000000      a      a01
a02 b03 c01 d30 e01 f61 g8000 h84      a02      a02      aLC 00958nam 22003371 450

```

図4. LC/MARC Communication format

format (図4) をフロッピーディスクか、あるいはハードディスクに出力可能である。これを直接 EB CDIC コードに変換して本体にのせ、拡張 COBOL で扱ってしまう方が簡便である。というのは、多くのパソコン用言語では扱うときにメモリー上の制約があるため、より煩雑なロジックを要求されるからである。しかしながら多くの CD-ROM の利用様態が、「パーソナルコンピュータでスタンドアロンによる運用」ということを勘案すると、PC 上で変換処理を行うほうがより汎用性の高い方法であると思われた。

今回手法としてはどこにでもある BASIC インタープリターを用いた。十分実用に耐えるものではあったが、あくまでこれ MS-DOS に付録のようにについてくる言語ソフトとしての一般性によるもので、プログラミングの労や実行処理速度を勘定に入れると、多少の投資をしても BASIC コンパイラや C 言語などを使用するほうが良いだろう。作業手順としては、

①コミュニケーションファイルからフィールド及びインディケータを切り出しそれに当該タグを付加した。これでパソコン上で扱いやすい一種の標準テキストファイルが、出来あがった (図5)。

```

#
001 87013805 /AC
008 $a870505s1987      nyua      j b      00110 eng
010 $a      87013805 /AC
020 $a0531103781
0500 $aTK7882.C56$bL36 1987
0820 $a004.5/3$219
10010$aLampton, Christopher.
24510$aCD ROMs /$cby Christopher Lampton.
2600 $aNew York :$bF. Watts,$c1987.
300 $a96 p. :$bill. ;$c23 cm.
4900 $aA First book
504 $aA CD-ROM library: p. 87-90.
500 $aIncludes index.
520 $aExplains how CD ROM and optical information storage work and discusses th
650 0$aCD-ROM$xJuvenile literature.
650 1$aCD-ROM.
952 $a002020
953 $a610618
954 $aZ320
955 $a0000000
956 $a0000000
957 $a
958 $a01
962 $a02$b03$c01$d30$e01$f61$g8000$h84
976 $a02
977 $a02
978 $aLC

```

図5. 変換後のテキストファイル

```
"87013805","0531103781","L a m p t o n , C h r i s t o p h e r . ","Lampton, C
hristopher. ","","CD ROMs /","","by Christopher Lampton. ","New York :","F. Watts
","","1987","96 p. :","23 cm. ","","Includes index. ","","CD-ROM. ","Juvenile lite
rature. ","","0020206","610618","Z320","0000000","0000000"," ","01","02","03","01
","30","01","61","8000","84","02","02","LC",""
```

図6. デリミテッド・アスキーファイル

②自館 MARC の個々のフィールドに、LC のタグを合わせた表を作成した。

③これをもとに、メインフレームの持つ自館 MARC のフィールドに合せたタグを割り付けていった。試験的に1バイト文字を2バイトに変換した。

1バイト文字はEBCDICに、2バイト文字は本体が日立であるため KEIS コードにそれぞれ変換した。

④他方で、前記テキストファイルから、市販データベースやパソコン用言語で扱いやすいレコード形式に変換した。

⑤デリミテッド・アスキー形式のファイルを作成した(図6)。

⑥PC上で各種帳票の出力を行なった。

1-2. プログラム作成

先ず、可変長データからテキスト形式のデータに変換するプログラムを組み、次に変換後に乗せるDBにあわせて異なった2つのプログラムを準備した。パソコン用には、タグとインディケータを読み込みながらサブフィールドごとに切り出し、デリミテッド・アスキーファイルに変換することとした。ホストDB用には、ホストDBが持っているバッチ入力用のformat形式にあうファイルを作った。

本学の場合、1レコードが270バイトの固定長ファイルである。バッチ入力口の無い場合は、本体側でもう1本プログラムを作る必要がある。

以下のプログラム作成のために周知のこととは思いますが、LC/MARC Communication format のレコード形式の概要に改めて触れておく。

一般に我々が知っている LC/MARC とは、9トラック・1600BPI のテープで供給されるもので、EBCDIC コード・標準ラベル付・可変長レコードで

構成されている。

LC/MARC の種類としては、単行本を収めた「Books All」、雑誌等を収めた「Serials」、楽譜などの音楽関係を収めた「Music」、地図を収めた「Map」、視聴覚資料を収めた「Visual Materials」等がある、また著者名典拠を収めた「Name Authorities」や件名典拠を収めた「Subject Authorities」が、それぞれMTで供給されている。

1レコードは図7に示すとおり、リーダー部・ディレクトリー部・データ部から成っている。リーダー部は、論理的レコード長とそのレコードの状態表示から構成されている。ディレクトリー部には、それぞれのフィールドNoとその位置と長さに関する情報が記述されている。データ部は、サブフィールドマーカがー1Fー、フィールドマーカがー1Eー、それにレコードの終わりがー1Dーでそれぞれ区切られている。これらを利用して1バイトずつに読み込んで行き、当該フィールドマーカとサブフィールドマーカを変換して、更にフィールド単位でー0D0Aーのり

Converted record (MARC FORMAT)

```
LEADER =
|00783cam 22003011 4500|

DIR =
|0010015000000008004100015010001900
0014125000120022426000400023630000
0011003589540009003699550012003789
3976000700460977000700467978000700

DATA =
| 62011180 /L|74021361953 nyu
BF121$b.H5 1962| $a150|10$aHilgar
o psychology.$cUnder the general e
aNew York,$bHarcourt, Brace $c|195
liography.| 0$aPsychology.| $a060
Total Records Read: 1 a02$b03
Total Records Converted: 1
```

図7. LC/MARC 構造

```

1  '-----< Copyright,1988 by Komatsu, Yasunobu >
2  '
35 OPEN "com.cp" FOR INPUT AS #1
36 OPEN "ccc" FOR OUTPUT AS #2
37 IF EOF(1) THEN 60
40 A$=INPUT$(1,1)
45 IF A$=CHR$(&H1F) THEN 57
50 IF A$=CHR$(&H1E) THEN 55
51 IF A$=CHR$(34) THEN 59
53 IF A$=CHR$(&H1D) THEN 56
54 PRINT #2,A$;:PRINT A$;:GOTO 37
55 PRINT #2,"":GOTO 37
56 D=ASC(A$)+6:E$=CHR$(D):PRINT #2,E$:GOTO 37
57 B=ASC(A$)+5:C$=CHR$(B)
58 PRINT #2,C$;:GOTO 37
59 PRINT #2,"":GOTO 37
60 CLOSE #1,#2
100 OPEN "CCC" FOR INPUT AS #1
101 OPEN "CCT" FOR OUTPUT AS #2:GOTO 110
106 IF EOF(1) THEN 600
107 ERASE C$,D$,E$,F$
110 A$=INPUT$(24,1)
113 PRINT A$
114 K=80
115 DIM C$(K),D$(K),E$(K),F$(K)
120 FOR I=0 TO K
220 C$(I)=INPUT$(3,1)
230 B$=LEFT$(C$(I),1)
250 IF B$=CHR$(&HD) THEN 320
260 D$(I)=INPUT$(4,1):F$(I)=INPUT$(5,1)
270 PRINT C$(K)
310 NEXT I
320 FOR J=0 TO K
330 LINE INPUT#1,E$(J)
333 E$=E$(J):E=LEN(E$):IF E>254 THEN 2000
335 K$=LEFT$(E$(J),1)
337 IF K$="#" GOTO 360
350 NEXT J
360 FOR H=0 TO I+1
370 PRINT #2,C$(H);E$(H)
390 NEXT H
400 GOTO 106
600 PRINT"
690 CLOSE #1,#2
695 KILL"ccc"
700 END

```

*** 終了! ***

図8. 変換用プログラム例

ターンコードを付与することとした。それにインディケータと配列を利用しタグ番号を付加したものが図5のファイルである。

先にも述べたとおりパソコン上でプログラムを組む時に注意すべき点は、できるだけメモリーに依存しないようにすることである。また、データの多目的利用を目論むなら、デバックがしやすい小さく平易なプログラムを途中から何本も枝別れさせて並列に組むと便

利である。フィールド長を計算してこのプログラムの逆を辿れば可変長 LC/MARC Communication format ができあがる (図8)。

1-3. JAPAN/MARC 及び日版 MARC CD-ROM 版のデータ形式

現在発売されている日本の CD-ROM 版 MARC のうち代表的な J-BISC (JAPAN/MARC, CD-

ROM)⁴⁾とCD-NOCS(日販MARC, CD-ROM)のデータ形式は、前記プログラムで作成されたデータと、インディケータ・タグNo等を除けばほぼ同じ形式となる。前節で述べたとおり、これら日本版CD-ROMデータも前出プログラムの逆を辿ることで可変長MARC format化することが出来る。

1-4. Diacritical mark (分音記号)の変換

今回対象とした資料には、英語以外の言語は少なかったが、一応、外字も変換してみた。ウムラウトや、アクサンテギュールといった外字は、BiblioFile CD-ROM側と日立・図書館システム側とがそれぞれ別の外字コードを持っているので、それに対応していくように変換した。

CD-ROM 接続パソコンでは、表示・印字とも出来る。メインフレーム側では、学術情報センター接続の専用目録端末とプリンターでなければ印字は勿論、表示もできない。これは、NACSIS 接続の各館と事情は同じである。

2. PCでのデータ利用

メインフレームのラインプリンターやCD-ROM接続のPCのプリンターで打ち出せない帳票を打ち出すため、市販DBやPC用の言語で扱いやすいファイルを作った。ラインプリンター・デュアルモードプリンター・タイプライター等で打ち出せない帳票とは、英文小文字と漢字の入ったリストや、2バイト数字に変換した図書用背ラベルの他(ラインプリンターは本体接続モードの場合小文字が打ち出せない)各種統計資料類である。

標準テキストファイルからデリミテッド・アスキーファイルに変換することにより、プリントアウトが可能となる。この形式のファイルは、市販DBでもPC用言語でも扱いやすい形式といえる¹⁾。こうした随時のリスト出力に対応出来るのがPCのメリットである。

パーソナルコンピュータだけで図書システムを構築している図書館も少なくないが、こうしたシステムでは他ファイルからの入力機能を持ったものが多い。市

販DBからデータの取り込みが出来たり、データのやりとりが出来ようになっているものも少なくない。たとえ、そうした入口の全くないシステムでも、同一OS(MS-DOS)上で動いているシステムならば、前記ファイルからわずかのプログラム作成によってデータの取り込みが可能となる。

3-1. オフコンへのデータ変換

まず、ホストへのデータ変換について解説する。本体はHITAC L-470Xで、OSはVOSO/ES Level 7.17。図書システムはLOOKS/U ver. 02.00である。

3.5インチ2HDのディスクを8インチのMS-DOSファイルに媒体変換後、本体のユーティリティを使ってEBCDICおよびKEISに変換する。この時点で8インチ2DDのLOOKS/U MARCがフロッピー上にできあがる。

多くのシステムがそうであるように、LOOKS/Uにはバッチ処理用入力口はあるのだが、J/MARCとTRC/MARCの受け口しかないため、直接ディスクのLOOKS/U MARCに相当するワークファイルに書き込みに行く。このワークを利用して多巻本や複本等の判定を行ない、書誌ファイルと図書ファイルに落とし込む。参考までに、LOOKS/U MARC項目と、今回乗せたLC/MARCのサブフィールドの対照表を示す(表1)。

この作業の過程でデータのアップロードには、乗せるデータとそれが入る器であるシステムとの、相対的関わりから決定する必要がある事を思い知らされた。一例ではあるがLOOKS/Uの場合、「ヨミ」のフィールドに入れたデータは、検索キーとして自動切り出しが行なわれる。しかし、著者名等のフィールドについては前方一致の検索キーになり、フィールドに余裕があっても1フィールド96バイト1名とされてしまう。件名についても、ワカチガキによるワード単位で切り出されるため、AACR2で見られるようなハイフンでつないだSubjectを入れても期待した検索結果は得られない。これらの問題を克服するために、いわゆる「Stop Word」を多数設定しておくのも一法であろう。

表1. 羽衣 LOOKS/U MARC 対照表

項番	項目名	属性	BIT	LC Tag	項番	項目名	属性	BIT	
1	和洋区分	X	2	976\$a	35	第二著者名	K	64	
2	シリーズ単区分	X	2	977\$a	36	同 (ヨミ)	X	64	700\$a
3	分類番号	X	7	957\$a	37	巻の第一著者名	K	96	
4	予備	X	3		38	同 (ヨミ)	X	96	(100\$a)
5	ISBN	X	13	020\$a	39	巻の第二著者名	K	64	
6	予備	X	3		40	同 (ヨミ)	X	64	(700\$a)
7	MARC区分	X	3	978\$a	41	件名1	K	32	
8	MARC・NO	X	12	010\$a	42	件名1 (ヨミ)	X	32	650\$a
9	予備	X	3		43	件名2	K	32	
10	出版地	K	10		44	件名2 (ヨミ)	X	32	600\$a
11	同 (ヨミ)	X	10	260\$a	45	内容注記	K	48	
12	出版社	K	20		46	内容注記 (ヨミ)	X	48	500\$a
13	同 (ヨミ)	X	20	260\$b	47	一般注記	K	48	
14	発売年月日	9	8	260\$c	48	一般注記 (ヨミ)	X	48	505\$a
15	巻次	9	5		49	予備	X	272	
16	版表示	K	20		50	図書番号	X	20	952\$a
17	同 (ヨミ)	X	20	250\$a	51	予備	X	3	
18	ページ数	X	9	300\$a	52	館コード	X	2	958\$a
19	大きさ	X	10	300\$c	53	受入年月日	9	6	953\$a
20	予備	X	10		54	価格 (定価)	9	7	955\$a
21	第一書名	K	96		55	請求記号	X	16	957\$a
22	同 (ヨミ)	X	96	245\$a	56	予備	X	3	
23	第二書名	K	96		57	購入業者コード	X	4	954\$a
24	同 (ヨミ)	X	96	245\$c	58	受入価格	9	7	956\$a
25	巻の第一書名	K	96		59	受入区分	X	2	962\$a
26	同 (ヨミ)	X	96	244\$a	60	管理区分	X	2	962\$b
27	巻の第二書名	K	96		61	資料区分	X	2	962\$c
28	同 (ヨミ)	X	96	249\$a	62	取扱区分	X	2	962\$d
29	副書名	K	96		63	配架区分	X	2	962\$e
30	副書名 (ヨミ)	X	96	245\$b	64	年度	9	2	962\$f
31	題書名	K	96		65	部署コード	X	4	962\$g
32	題書名 (ヨミ)	X	96	440\$a (490\$a)	66	頁目コード	X	2	962\$h
33	第一著者名	K	96		67	予備	X	44	
34	同 (ヨミ)	X	96	100\$a	68				

3-2. CD-ROM 版 MARC・書誌ユーティリティ・ローカルDB

データコンバートを考える際重要な問題は2つある。第一は乗せるデータの性質を知っておくことであり、第二は受け手側のシステムの性質を知っておくことである。今回はCD-ROM版LC/MARCのデータをコンバートしたわけだが、図書システムとデータの関係は、さながらローカルDBと書誌ユーティリティの

関わりを連想させる。

汎用機・オフコン・パソコンを問わず一般的な図書システムは、漢字とそのヨミから構成されている。これらは、JAPAN/MARC準拠、或いは類似のformat構造を持っている。表1を見て頂ければわかるように、タイトルにあたる書名1ヨミは96バイトで、これを漢字にKEISE変換するならば98/2=48バイトとなり、桁溢れが多く表示のみで意味がない。そこ

で検索機能を考慮し、書名1及び2のヨミを一つのフィールドと見なして TITLE 245 \$a および責任表示に当たる 245 \$c を合わせ出力してみた。これにより 245 \$a の桁溢れはほぼなくなった。むしろ、Imprint にあたる出版地・出版社などの出版事項のオーバーフローが著しいものとなった。

短絡的にローカルDB = MARC format という立場を取るならば、ローカルDBのformatを改造する必要が生じてくることになる。しかしこの改造を手がけるには、オンライン書誌ユーティリティやローカルDBそれにCD-ROMメディアといった3つの有機的関係の確立を考案することが先決であろう。この問題は、VI-2で再説する。

IV. メインフレームとCD-ROMとの可能性

PC単体でCD-ROMが動く一方、メインフレームを中心にオンラインでハウスキーピング業務を行い、ローカルデータを結合した書誌データを月2回、バッチで乗せるというのが本学の現状である。あるいは前記の方法で、メインフレームへ可変長データであるCommunication MARC formatのまま転送し、ホスト上で処理することも考えられる。いずれにしてもエミュレータを使うかフロッピー渡しりのバッチ処理となる。

1. コミュニケーション手順

今回採用した手順とは異なり、CD-ROMと接続したPCを、メインフレームに直結して取り込んだデータをメインフレームの複数端末で処理していくという方法もある。こうした発想の多くはローカルデータの効率的処理を目的としたものである。

BiblioFileにはコミュニケーション機能(図9)があるため、ホストコンピュータや他のコンピュータへRS-232Cをつかって直接データを転送することができる。通信形式には、CLSI protocol・X-ON X-OFF protocol・ALIS protocol等が用意されている。ボーレートは300~9600ボーから選べ、語長は7~8bitで、Stop bitとParityを決定して

Sect 5: Communication Setup		BIBLIOFILE
1. Communication format:	MARC	Catalog
2. Communication port:	COM1:	Production
3. Baud rate:	C_1200	Configuration
4. Word length:	8_BIT	
5. Stop bits:	1_STOP	
6. Parity:	C_NONE	
7. Path to Comm Output File		
comm:		
8. Overlay code:	BI	
X. Return to last menu		
Enter Selection:		

図9. コミュニケーション機能

Communication Portに接続すれば通信可能となる。

2. メインフレームからI/Oのメモリー渡しの可能性

これまでに発表されたCD-ROMシステムの多くは、“出口”のインターフェース(ダウンロード等)がそれなりに確保されているものでも、“入口”のインターフェース(キー入力すべき検索キーをファイルとして溜めたものから、自動的にヒットさせていく等)は、PCからのキー入力のみ頼ってきた。つまり検索キーは、常に一人のオペレータによってキーボードより入力されねばならなかった。BiblioFile v3.19は、発注データを自動検索入力用データとしてTransaction fileに“入口”を開放している⁵⁾。技術的に、この機能を拡張することは、難しい問題ではない。またv4.03では、インデックスディスク以外にヒットしたキーを蓄積出来るキューファイルの一覧が可能である。さらに新しいバージョンでは、Consideration fileとして他のダウンロードデータを検索用キーとして取り込める。

例えば、メインフレーム接続したCD-ROMに検索用キー項目を蓄積し、それを一括してメモリー渡してPCに送る。PCでは送られてきたデータにII-3で述べたローカルデータの不変項目と可変項目を突き合わせ、ヒットしたものは結合する。

この場合の可変項目は、検索キーと共に転送されてくるものとする。完成したデータは、メモリー渡してそのままメインフレームに取り込まれていく。ヒット

しなかったものは、不変ローカルデータと可変キー項目を含むローカルデータに検索キーとなった項目のみ自動入力されたものが結合され、別ファイルに分けられ、後でオリジナル入力用ファイルとして修正をまつ。修正はメインフレーム側端末でもPC端末でも可能となるだろう。

使用する端末の入力用エディターの性能と端末の数、作業効率に比例することになる。修正の済んだものからメインフレームDBに、ヒット分と同じように転送されていく。

以上は、あくまでもメインフレームを中心に据えたシステムを想定している。これには技術的な問題よりも、CD-ROMを取り巻く法的問題のほうがより障害となるであろう。現状では単体のPCにCD-ROMが接続されているので、特定のユーザーが常に見えている。ところが、メインフレームとの接続によって、特定出来るユーザーが見えにくくなることは否めない。その結果発生してくる利用規定上の問題は、十分に議論される必要があろう。

V. NETWORKの可能性

CD-ROMを利用したデータ処理には、メインフレームを利用する形態が最終目標になるのだろうか。ここでPCのLAN(図10)について考えてみたい。

DB型CD-ROMは、データ量の増加と検索機能

Sect 6: CD-ROM Configuration		BIBLIOFILE
I/O Port Address for connected CD-ROM reader(s): (Do NOT change these fields unless instructed!)		Catalog Production Configuration
1. Drive 1:	LOC_CD H_1503 300	
2. Drive 2:	NO_CD	
Possible values and value meanings:		
1. NET_CD1 - 1st CD-ROM on network		
2. NET_CD2 - 2nd CD-ROM on network		
3. NET_CD3 - 3rd CD-ROM on network		
4. NET_CD4 - 4th CD-ROM on network		
5. NET_CD5 - 5th CD-ROM on network		
6. NET_CD6 - 6th CD-ROM on network		
7. NET_CD7 - 7th CD-ROM on network		
8. NET_CD8 - 8th CD-ROM on network		
Pressing 'x' or <RETURN> will retain current value. Enter selection:		

図10. CD-ROM Network 機能

の充実から、いわゆる「マルチ・ボリュームセット」(多巻にわたるCD)が今後増えるだろう。その処理方法の一つとしてCD-ROMドライブの拡張機能がある。インデックス部・データ部の入ったCD-ROMをそれぞれ収めた複数ドライブを連結して1台のPCに接続し、全体として1枚のDBにアクセスしているような操作性を実現しようとするものである。

このエンハンスド接続した1台のPCをサーバーとして複数のPCからアクセスすれば、マルチアクセスを行なっているように使用できる。

BiblioFileの場合、検索キー入力で最初のCD-ROMへのアクセスがなされる。ヒットリスト表示で、接続してあるPC上のメモリーにヒットしたインデックスが貯えられていく。更に、当該書誌についているNoを検索するときに2度目のアクセスがなされる。それぞれのアクセスは、スタンドアロンで利用している限り極めて速いものであった。

CDの利用がスタンドアロンから発展して、1台のCD-ROMに複数のワークステーションからアクセスすることが可能となるならば、その経済性は飛躍的に増大する。それだけではなく、メインフレームを中心としたシステム体制、すなわちデータベースマネジメントの在り方にすら影響を与えかねない。

VI. むすび

これまでに、約1,000冊のアップロードを終えた。この結果によって実用性を結論してみる。

アップロードしたデータからわかることは、1930～50年の出版物であるため、直接ヒットした件数が、出版年・版表示をはじめとする書誌修正を含めないと60%を超えないということである。また一方、全くヒットしなかったものについては、LC/MARC準拠のオリジナル入力を行なった。これは、将来の書誌データ標準化を考慮してホストへの直接入力を避けたものである。

運用形態は、今のところスタンドアロンのPCであるため、オペレーターは一人である。

明らかになった問題点は、

- ①ローカルデータの作成・入力・追加の方法は、CD-ROM 検索用ソフトとユーザーが独自に組むソフトのいずれかに依存する。
- ②オリジナル入力を含む書誌そのものを作成・変更する機能は、遡及入力の場合特に有効である。
- ③1台のCD-ROM に対して複数の端末で作業を行えば、乗数分の能率が得られるほどにアクセスは速い。換言すれば、CD-ROM を操作するソフトの側で、NETWORK 等に対応した機能を必要とする。

1. ローカル館内部での CD-ROM の利用

これまでのバッチ処理は、遡及入力をその目的としてきた。遡及入力には、カレントなデータを必要としないためバッチ処理は有効であった。更に発注・受入の段階から CD-ROM のデータを投入することによって「入口」の帳票である発注票・物品請求命令書等にも利用・反映させ、全体的なデータ処理をはかると共に、そうすることで作業の重点を現在の整理・目録中心から発注・受入中心へ移行し、整理受入業務の流れを円滑にする事が可能であろう。そうした利用を前提とした BiblioFile の場合、出版前の CIP データを含むため問題はなかったが、カレントな利用を前提とする場合 CD-ROM メディアデータのタイムラグは許されない。

2. 相互利用のためのデータの独立

データ変換作業は、ハウスキーピング業務にどう役立てることが出来るかを、勘案して論を進めてきた。そのためトータル管理システムである本学の HOST 図書 DB (ローカル DB) に変換すると同時に、背ラベル・冊子体目録リスト・統計資料等の整理業務に役立つ帳票を PC から直接出力出来るようデータ加工を行った。

自館内部でのデータ加工は、遡及入力と発注に役立てることで確立されるが、将来、共同目録作業・相互利用にそのままメインフレーム DB (ローカル DB) の利用を考えるならばⅢ-3-2 で考察したように、図書システムのなんらかの format 改造を図らねばならない。はたして個々の図書館の状況に対応したメイ

ンフレーム図書 DB にそうした多面的機能を負わせる事が出来るかどうか、現状においてはやや疑問である。

むしろ、MARC の CD-ROM ならばデータとしての標準性からもより多目的利用に対応出来るであろう。このためには遡及入力の場合、未ヒット分の処理を考えると今回行ったような CD-ROM からのヒット・書誌修正と同一の MARC format でのオリジナル入力、およびその出力が必要となる。この場合のデータの流れは、

CD-ROM メディア→メインフレーム図書 DB (オリジナル入力)→オンライン共同目録

という流れではなく、

自館 HOST 図書 DB←オンライン←オンライン共同目録

 (オリジナル入力)

となり、自館図書 DB からオンライン共同目録のためのデータは独立する。

データは、それが入る器であるシステムに規定される性質がある。本論で述べたのは、データ処理の舞台となる自館図書システムとソースデータの作成に二度手間の入力を行うことなく、またメインフレーム図書 DB にデータを直接入力して自館システムのみデータに従属させることもなく、発注の時点から役立つデータと遡及入力を統一データ形式で分散入力し、その多目的利用を図ることを目的としている。言い変えるならば、データマネージメントをシステムから独立させ、且つそれを有機的に結合させようという提案である。

1988 年夏の American Library Association 総会で米議会図書館が LC/MARC の全面 CD-ROM 化計画⁵⁾を打ち出したが、このメディアは今後利用法を検討する必要に迫られよう。

謝 辞

最後になるが、今回の HOST 図書 DB へのデータ変換作業を承諾してくれた、布川事務長に感謝したい。

引用文献

- 1) 朝倉秀三 パーソナルコンピュータによる JAPAN/MARC の利用に関する研究。情報管理

1987 ; 29 (11) : 939-50.

参考文献

- 2) User's Guide for BiblioFile Catalog Maintenance System. Washington, D. C. : The Library Corporation, 1987.
- 3) Catalog production system user guide 3. 12 Revision. Washington, D. C. : The Library Corporation, 1988.
- 4) 鶴田真也 Japan MARC CD-ROM 研究開発特別委員会報告. 図書館雑誌 1988 ; 81 (9) : 564-5.
- 5) 布川嘉佑, 小松泰信 LC MARC CD-ROM Edition (The BiblioFile) について. 情報の科学と技術 1988 ; 38 (6) : 289-99.
- 6) LC MARC Subjects / Reference Manual. Washington, D. C. : Library of Congress, 1988.
- 1) Crawford W. MARC for library use : understanding the USMARC formats. New York : Knowledge Industry Publication, 1984.
- 2) 藤原譲 CD-ROM ミニ・シリーズ総括. 情報管理 1988 ; 30 (12) : 1177-80.
- 3) MARC formats for bibliographic data. Washington, D. C. : Library of Congress, Automated Systems Office, 1987.
- 4) 目録システム利用マニュアル. 学術情報センター, 1986.
- 5) 目録所在情報サービス利用の手引. 文献情報センター, 1985.