

東大寺文書データについて

照井武彦

はじめに

東大寺文書目録のデータベースをつくったので報告する。同目録をもとに、最初につくったデータファイルは、目録表記の細部にわたって表現するものとした。即ち、傍注や割り書きをも記号をきめてデータに含めるようにした。これらは目録の表記を、適当な出力用ソフトがあれば、細部まで再現するに十分な情報を持っている。

オンライン検索用データベースは、そのようなデータファイルから出発して、記号や用字を自動的に整理し、形式を変換することによって、容易に搭載用原ファイルを得ることができ、実際にそのようにしてオンライン検索用データベースに搭載された。

以下に、これらのことを具体的に詳しく述べ、関連する諸問題について若干の議論を行う。

1 東大寺文書目録について

東大寺文書目録^[1]は6巻から成り、そのうち第一巻から第五巻までに目録本体を載せ、第六巻の大部分は編年目録で、他巻と編成を異にしている。第五巻までの各巻は、口絵図版、凡例、目次、目録本体、あとがきの順に構成されている（A5版）。その諸元はつぎの通りである。

	頁数	発行年次	所 載
第一巻	488	昭54	東大寺文書 第1部（寺領部）第1～第23
二	384	55	“ “（ “ ）第24～第26
			“ 第2部（寺法部）
三	360	56	“ 第3部（文書部）第1～第9
四	387	57	“ “（ “ ）第10～第12
			“ 第4部～第6部
五	487	58	“ 第7部～第142部
			宝庫文書、薬師院文書
六	—	59	編年目録、指定品目録

今回のデータ作成に当り、対象としたのは第一巻から第五巻までであるが、編年目録をみると、これらと並んで第六巻所載のものが若干ある。

また、東大寺以外にある文書について田中稔氏が調査した結果が仮目録の形になって居り、その写しを筆者が預っている。それは調査途中なのか、終了したのか、よくわからない。

2 データ化の考察

データ入力を開始するまでには、かなり時間を要した。それは先行する国宝重要文化財総合目録（美術工芸品編）データ作成の経験を活かしたく、両者の並行を避けることにしたからである。同目録データについては既報告^[3]で詳しく述べ、データ化の困難性とその解決についても触れているが、東大寺文書目録にも特有の困難性があった。しかし、共通の問題で先の経験が役立った事も少なくない。

特有の問題としては、傍注とそれに使われている記号、および文書番号の親子関係などを挙げることができる。

一方、共通の問題としては、旧字体漢字など第2水準漢字の多発と、データ項目の選別（種類決定）がある。前者は経験が良く活きた例、後者は活きなかった例である。

なお、読みガナは今回も付さなかった。この問題についてはすでに報告済み^{[2][3]}であるが、公開中のデータベースは漢字検索によって居り、特に問題視する状況ではなくなったと考えられる。

データ構造については、凡例にデータ項目となるべき事項がつぎのように列挙されているので、ほぼこれに従うことにした。データ入力の途中で、「表紙」項目を設けたいとの希望が出されたので、これを加えることにした。

- | | |
|------------|------------------|
| (1) 番号 | (10) 本文の首部 |
| (2) 文書名 | (11) 本文の末尾 |
| (3) 欠損 | (12) 差出 |
| (4) 日附 | (13) 充所 |
| (5) 寸法 | (14) 表書（ウハ書） |
| (6) 紙数 | (15) 切封帯（ある場合のみ） |
| (7) 時代 | (16) 封墨引 |
| (8) 形状 | (17) 特記すべき事項 |
| (9) 端裏書・端書 | (18) 脚注（主として庄園名） |

このように、1件の中には同種項目の入れ子構造のようなもの⁽¹⁾がなく、平坦な構造になっているので対処は容易である。しかし、一段上の構造をみると、文書番号に親子関係があり、しかも番号と親子関係の対応も一様ではない。その対処の仕方は難しく、オンラインデータベースに搭載するデータ構造としては一応の解決をみているが、他の応用に対して万全とは言えない現状である。これもひとつの入れ子構造である。多様な入れ子構造を容易に、かつ自然にデータ化できるような技法は、この分野ではどうしても必要と考えられ、この方面の技術開発が望まれるところである。⁽²⁾

データ設計において、今回、特に試みたのは、目録書式情報の取り込みである。そのため、多数の記号を利用した。ここで書式と言ったのは、主として傍注と割り書きである。データは書式

を表わす記号が各所に挿入された文字列として構成され、1項目が1行につくられる。

こうすることによって、データファイルには、原目録の書式を再現するのに十分な情報を含むことになる。今回のデータから、そのような印刷を自動的に行うソフトウェアはまだないが、周知のように電子出版が普及しつつある現在、その方向に研究を一步ふみだしておく必要がある。さらに言えば、書式情報の分解も興味ある課題である。^[4]

ここで、着手以来、現在までの経過を表1に示す。

表1 作業の年次経過

年度	西暦	事 項
昭和61	(1986)	試入力(第一部美濃国大井庄) 70頁
〃 62	(1987)	入力
〃 63	(1988)	入力
平成1	(1989)	校正
〃 2	(1990)	校正
〃 3	(1991)	校正, 公開準備
〃 4	(1992)	校正, オンライン公開(10月)

また、付録1～5に、原文書から目録、データにいたる各段階を、同一文書について表示し理解の助けとする。

3 データファイル

目録を入力して得られるデータファイルの利用目的がいろいろあることは前節で述べた。その全部を入力で段階で予想するのは困難である。データ設計に反映させるのは、オンラインデータベース(具体的には HITAC VOS-3 の ORION システムにのせる事になる)と、目録書式の再

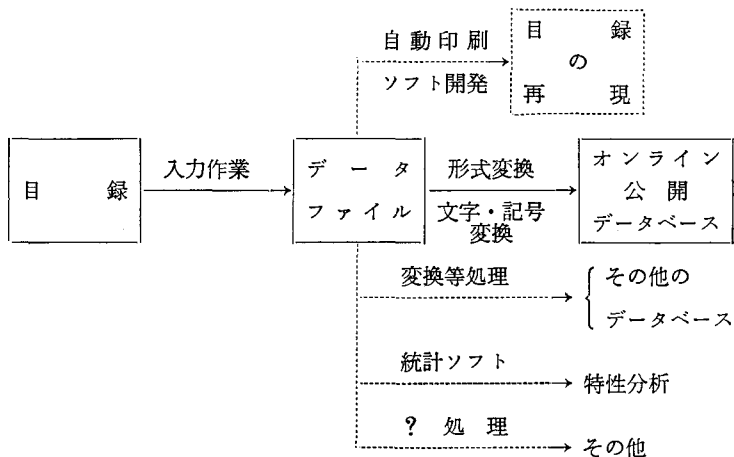


図1 データファイルの応用

現とのふたつにしておけば、この先かなりの応用範囲をカバーできるであろう。以上の考え方をまとめて図1に示す。

さて、データファイルの諸元はつぎのようである。

- レコード特性 可変長ブロック化
- 最大レコード長 255バイト
- ブロック長 5104バイト

これは、現在の歴博ホストコンピュータ HITAC-660, VOS-3 OS を前提にしている。

1レコードの構成は、左から1バイト固定長部、2バイト可変長部から成る。

- 1バイト固定長部

1バイト 空白又は* (次レコードがこのレコードの継続であることを表わす)

7バイト レコード番号

1バイト 空白

4バイト 目録の巻頁を1, 3バイトで表わす。

1バイト 空白

6バイト 文書の親番号を表わす

1バイト 空白

6バイト 子番号, 孫番号, 紙背番号

1バイト 空白

2バイト 項目番号

1バイト 空白

- 2バイト可変長部

各項目の内容を2バイト可変長部に漢字データとして表わす。途中に各種記号を含む。その全体の両端を漢字 in, 漢字 out コードで囲む。

レコードの継続は3~4に及ぶものがある。それは割り書きを1行に並べかえることと、多数の記号の追加が重なって起るからである。

目録で明朝体の数字は子番号, ()つき洋数字は孫番号, 紙背は原則として親番号の直属としたが、紙背番号の例外が2箇所ある。

データファイル上、1件の定義は確定していない(固定長部に件番号欄を置けば確定する)。1件の定義は3通り考えられる。その一は親番号のみによるもので、子、孫が多数つくものは非常に大きくなる。その二は親、子、孫、紙背を対等に各1件とするものである。下位からみて、直属の上位各件が離れて見にくくなる。その三は、その二と似ているが、最下位を単位とし、その都度上位のものを先行付加して1件とするものである。上位のものが多数重複し、全体が大きくなる。後で述べるように、オンラインデータベースでは、その二を採用している。

つぎにデータ項目であるが、これは目録の項目を、凡例記載の番号を含めてほぼそのまま用い

ることとした。ただし、「表紙」項目も必要とのことで、それと備考、地名を設けたことが主な変更点である。

表2 データ項目番号と項目名

番号	項目名	番号	項目名
01	文書番号	11	本文の末尾
02	文書名	12	差出
03	欠損	13	充所
04	日付	14	表書
05	寸法（表示のみ）	15	切封帯
06	紙数（表示のみ）	16	封墨引
07	時代	17	特記事項
08	形状	29	表紙
09	端裏書	30	備考
10	本文の首部	31	地名

項目の並び順、項目の内容の用字は目録になるべく忠実にしたがうようにした。目録で使われている記号のうち、JIS（日本工業規格）C 6226 非漢字で表わすことのできるものはそのまま用いた。これを第1類の記号と呼ぶことにする。非漢字で表わせない記号は非漢字から選んで置きかえた。これを第2類と呼ぶ。さらに、書式を表わす記号を追加した。これが第3類である。つぎに、これらの記号の例を挙げる。

第1類	第2類	第3類
「0154	々→〃0123	割書右△0204と▲0205
×0163	〈→∴0172	小文字≥0170と≤0171
/0131	ろ→Φ0181	

ここに4桁の数字は各記号の区点番号である。

傍注は、最終的に読み下しが自然になるように挿入位置をきめたが、入力時に文字の相対位置から機械的に挿入したため、校正（読み下しの判断）とデータファイル修正（挿入位置の調整）は特に難航した。しかし、その結果、目録につけられていた記号に助けられて、かなり読みやすくなったように思われる。また、入力時にとまどったのは、原文書で言うところの線を引いて消し、訂正したような個所の記号表現であった。これは訂正前後を3種の記号で囲うことで対処した。第3類の記号があることにより、例えば傍注のみ取り出す、又は全部外すと言ったことが可能となる。

校正に当っては、同一項目を抜き出したリストが有効であった。異項目のデータの混入が一目でわかるし、その他の誤りも不規則なところを目で追うことで発見が容易であった。01文書番号

について抜き出したリストは、校正のほかに、親、子、孫のような番号構造を見定め、考察するのに役立った。

データファイルに関する処理環境は、すべて歴博電子計算機システムの汎用ホストコンピュータによった。

計算機	HITAC 660
オペレーティングシステム	VOS-3
テキストエディタ	ASPEN
プログラミング言語	VOS-3 C言語

ASPEN は、校正リストによるデータ修正に用いたが、傍注の挿入位置を調整するなどの細かい修正を自然に行うことができず、また日本語への対応が充分でないところがある。しかし、大きなファイルを分割せずに、検索も割合に高速で、この点便利であった。後から考えると、パソコンにダウンロードして dBASE のようなデータベースソフトを道具にしてデータ修正することもありえたと思うが、この場合にはファイルの分割やダウン・アップロードの繰返しなどに手間取るのを無視できないように思われる。

ASPEN はまた、データファイルの状態でいわゆる一字検索をするのに有用である。図 2 に、その様子を示す。「謹言」を検索するには、コマンド／謹言／を実行させると（上段画面）、画面の中ほどに最初の検索結果が現われる（下段画面）。続けて次々に検索することができる。

C 言語は、可変長レコードのデータを簡単に扱えるので使用してみた。

4 オンラインデータベース検索システム

歴博オンラインデータベースは、平成 2 年 4 月にサービスを開始し、この東大寺文書目録データベースはその 5 番目に当る。平成 4 年 10 月にサービスをはじめた。^[7]

オンラインデータベースシステムは、つぎの通りである。

ホスト計算機	HITAC 660
オペレーティングシステム	VOS-3
検索システム	ORION
館内検索端末	3020, 3010
館外検索端末(1)	学術情報ネット
同 上 (2)	電話回線

(1)は各大学大型計算機センター経由で全国各地の大学と接続されて居り、各大学の端末で検索することができる。特に大計センターとは直接接続であって、センター内端末によると接続操作

```

EE ***** D S名 ('REFE900.TOU87.VB') *****
目盛行  ---+---1---+---2---+---3---+---4---+---5---+---6---+---7---
***** ** データ行の先頭 **
000100 0000001 1003 101001 000000 01 【一】
000110 0000002 1003 101001 000000 02 【黒田庄庄官諸職諸名宿直免注文】
000120 0000003 1003 101001 000000 04 【寛喜三年九月 日】
000130 0000004 1003 101001 000000 05 【二九・三×四二・三】
000140 0000005 1003 101001 000000 06 【一紙】
000150 0000006 1003 101001 000000 07 【鎌倉中期、】
000160 0000007 1003 101001 000000 08 【豎紙、】
000170 0000008 1003 101001 000000 09 【(端裏書)「黒田庄内宿直免注文」】
000180 0000009 1003 101001 000000 10 【(書出)「下司各〃十間 くもん各〃十間】
000190 0000010 1003 101001 000000 11 【(書止)「右注進状如件」】
000200 0000010 1003 101001 000000 31 【伊賀国黒田庄】
000210 0000011 1003 101002 000000 01 【二】
000220 0000012 1003 101002 000000 02 【黒田庄庄官百姓等請文】
000230 0000013 1003 101002 000000 04 【(年未詳)六月廿三日】
000240 0000014 1003 101002 000000 05 【三〇・二×五二・〇】
000250 0000015 1003 101002 000000 06 【一紙】
000260 0000016 1003 101002 000000 07 【平安院政期、】
000270 0000017 1003 101002 000000 08 【豎紙、】
000280 0000018 1003 101002 000000 09 【(端裏書)「黒田」】
ページマップ 1+ . . . . . 3450 現在ページ ( 1 )
コマンド ( / 謹言 / )

```

↑
検索文字列

この行が強調され「謹」の字
にカーソルが移動する

```

EE ***** D S名 ('REFE900.TOU87.VB') *****
目盛行  ---+---1---+---2---+---3---+---4---+---5---+---6---+---7---
000200 0000010 1003 101001 000000 31 【伊賀国黒田庄】
000210 0000011 1003 101002 000000 01 【二】
000220 0000012 1003 101002 000000 02 【黒田庄庄官百姓等請文】
000230 0000013 1003 101002 000000 04 【(年未詳)六月廿三日】
000240 0000014 1003 101002 000000 05 【三〇・二×五二・〇】
000250 0000015 1003 101002 000000 06 【一紙】
000260 0000016 1003 101002 000000 07 【平安院政期、】
000270 0000017 1003 101002 000000 08 【豎紙、】
000280 0000018 1003 101002 000000 09 【(端裏書)「黒田」】
000290 0000019 1003 101002 000000 10 【(書出)「此官省符問御下文御庄委細令披】
000300 0000020 1003 101002 000000 11 【(書止)「恐〃謹言」】 ←
000310 0000021 1003 101002 000000 12 【(差出)「黒田□□□〃〇【庄〃官】●〃】
000320 0000021 1003 101002 000000 31 【伊賀国黒田庄】
000330 0000022 1003 101003 000000 01 【三】
000340 0000023 1003 101003 000000 02 【東大寺三綱等申状案】
000350 0000024 1003 101003 000000 04 【弘安九年十一月 日】
000360 0000025 1003 101003 000000 05 【三三・八×一〇九・二】
000370 0000026 1003 101003 000000 06 【三紙】
000380 0000027 1003 101003 000000 07 【鎌倉中期、】
000390 0000028 1003 101003 000000 08 【続紙、】
ページマップ 1+ . . . . . 3450 現在ページ ( 1 )
コマンド (

```

図2 ASPEN画面エディタによる「謹言」の検索

```

AU>M ----- 起動コマンド
*****
*                                     情報処理係 (TEL 225)
*
* ■運用時間は、次のとおりです。
*   月～金曜日   9：00～17：30
* ■公開データベースの利用時間は次のとおりです。
*   月～金曜日   9：30～16：30
*   但し、偶数月の第2金曜日の午後は、定期保守のためサービス
*   を休止します。
* ■図書検索サービスは、毎週木曜日午後は運用を休止します。
* ■10月1日より東大寺文書目録データベースを公開します
*   ので、ご利用ください。マニュアルについては後日発送します。
*****
ORION 05-0?
    
```

このデータベースは、奈良国立文化財研究所編「東大寺文書目録第一～五巻」(同所発行)に基づき、当博物館で作成されたものです。原目録の著作権は奈良国立文化財研究所、データベースの著作権は当博物館が所有しております。データベースの作成者は故・田中稔(前情報資料研究部長)と照井武彦(同研究部教授)です。内容、使用方法等のお問い合わせは下記にお願いします。

国立歴史民俗博物館 代表電話 043-486-0123
 情報資料研究部 照井武彦 (内線 407)
 資料課 情報処理係 (内線 225)

```

* 11204 1/ AL:1 ----- 初期全件集合
コマンドを入力して下さい。
  2/D FOR 1 ----- 第1件を表示せよ
----- ( 1) -----
    
```

```

文書番号 / P01 : 1-1-1 (1003) --
文書名   / P02 : 黒田庄庄官諸職諸名宿直免注文
日付     / P04 : 寛喜三年九月 日
寸法     / P05 : 29.3 x 42.3
紙数     / P06 : 一紙
時代     / P07 : 鎌倉中期
形式     / P08 : 縦紙
端裏書   / P09 : 「黒田庄内宿直免注文」
本文の首部 / P10 : (書出) 「下司各々十間 くもん各々十間」
本文の末尾 / P11 : (書止) 「右注進状如件」
地名     / P31 : 伊賀国黒田庄
コマンドを入力して下さい。
  2/?? ----- 項目表を表示せよ
    
```

項目	プレフィクス	属性	表示項目
文書番号	P01	EBCDIK	BAN
文書名	P02	漢字	
欠損	P03	漢字	
日付	P04	漢字	
時代	P07	漢字	
形状	P08	漢字	

図3 オンラインデータベース検索

端裏書	P09	漢	字
本文の首部	P10	漢	字
本文の末尾	P11	漢	字
差出	P12	漢	字
充所	P13	漢	字
表書	P14	漢	字
切封帯	P15	漢	字
封墨引	P16	漢	字
特記事項	P17	漢	字
表紙	P29	漢	字
備考	P30	漢	字
地名	P31	漢	字
多項目	P00	漢	字

2/F P11:*謹言 ----- 「謹言」の検索(後方一致)

* 1488 2/ P11:*謹言 (178 ターム 連結)

コマンドを入力して下さい。

3/LI ----- 検索履歴を表示せよ

件数 行番号 要求

* 11204 1/ AL:1

* 1488 2/ P11:*謹言 (178 ターム 連結)

コマンドを入力して下さい。

3/D=2 FOR 1-2 ----- 検索結果の第1～2件目を

表示せよ

----- (1) -----

文書番号 / P01 : 1 - 1 - 2 (1 0 0 3) 二

文書名 / P02 : 黒田庄庄官百姓等請文

日付 / P04 : (年未詳) 六月廿三日

寸法 / P05 : 3 0 . 2 × 5 2 . 0

紙数 / P06 : 一紙

時代 / P07 : 平安院政期

形式 / P08 : 豎紙

端裏書 / P09 : 「黒田」

本文の首部/ P10 : (書出) 「此官省符間御下文御庄委細令披露候之処」

本文の末尾/ P11 : (書止) 「恐々謹言」

差出 / P13 : 「黒田□□□□ || [庄々官] || 百姓等 || 上 || 」 || (日下) ||

地名 / P31 : 伊賀国黒田庄

----- (2) -----

親子 : ひ孫

文書番号 / P01 : 1 - 1 - 3 - - - 1 (1 0 0 3) [紙背一]

文書名 / P02 : 東大寺衆徒等重申状案

日付 / P04 : (年月日未詳)

寸法 / P05 : 3 3 . 8 × 4 7 . 8

紙数 / P06 : 一紙

時代 / P07 : 鎌倉中期

形式 / P08 : 続紙

端裏書 / P09 : 「寺解并院宣案」

本文の首部/ P10 : (書出) 「東大寺衆徒等重誠惶誠恐謹言」

本文の末尾/ P11 : (書止) 「仍重誠惶誠恐謹言」

備考 / P30 : 院宣案ナシ

: 黒田庄悪党

地名 / P31 : 伊賀国黒田庄

コマンドを入力して下さい。

3/Q ----- 検索システム終了

が容易である。(2)はパソコンを端末エミュレータによって接続するものである。^[6]

基本的には漢字検索によっているが、そのためのかな漢字変換は各利用者端末のそれを用いる。⁽⁵⁾
ホスト側のかな漢字変換は全く利用しない。

検索の簡単な例を図3に示す。先程と同じく「謹言」(ここでは後方一致による)を検索している。

歴博オンラインデータベースの検索方式にコマンド方式とメニュー方式のふたつがある。両方可能なものもあるが、東大寺文書目録については現在のところコマンド方式のみ可能である。^[7] コマンド方式はやゝ難しい感じを与えるので、検索の手引きの内容でそれを補うようにしている。⁽⁶⁾

データファイルから検索システム ORION への搭載手順はつぎの通りである。

(1) データファイル中の記号を置換する。

第3類の記号を全部 〓0134に置換する。また、例えば、「同じく」の意味で使い分けられていたク0123と々0125について、前者を後者に置換し、統一する。

(2) 旧字体漢字を新字体に置換する。⁽⁷⁾

(3) 以上により、搭載用原ファイルを得る。

(4) 原ファイルを、ORION 搭載のための歴博標準形式に変換する。

(5) ORION 搭載仕様に合うよう項目順序の調整等を行う。

ORION の中で索引が生成されるが、その指定に当って、事実上いわゆる一字検索を行えるような工夫をしたのが特徴である。ORION はもともと前方一致と後方一致検索は可能であるが、ここに言う一字検索は、漢字1字以上の中間一致検索を意味するので、そのままでは対応できない。そこで、通常の検索語のほか、これから頭を1字欠く語、2字欠く語をつくる。これらに対して検索のときに前方一致を指定すれば、中間一致と同じ結果が得られる。索引語数は、これを行わない場合に比べて数倍に大きくなるが、やむを得ないものとする。

この考え方は、単語分割(分ち書きとも言う)とは異なるものと言える。単語分割は、分割された語がそれぞれの意味を持つものとして取り出す操作である。⁽⁸⁾ 他方、上の場合には意味に関係ない「文字および文字列」として扱う。ある語の後半と次の語の前半が続いた文字列でもかまわない。その気になれば、句読点などの記号さえも検索の対象とすることができる。⁽⁹⁾

本節の最後に1件の定義について述べる。ORION に搭載するに当り、親、子、孫、紙背を各1件とすることは前にも述べた。ORION ではこの方式しか採ることができない。親、子等で1件とすると、同じ項目番号が親にも子にも現われるが、これが許されないからである。検索結果が子だけ表示され、その親が見えないと不便なので、いろいろ検討した結果 EXPAND コマンドを使い、その親に属する子等の全体を別の検索集合につくれるようにした。

以上、すでに公開されているので、御使用いただき、使い勝手やデータの誤り等御指摘いただければ幸いである。⁽¹⁰⁾

5 検索システム論

ここに使用されている検索システム ORION は、IBM 社の STAIRS を祖とするもので、広義のデータベースシステムの中では最も歴史の古いものの一つである。この報告で検索システムと言ってデータベースと言わなかったのは、データベースは一般の利用者によって日々更新され得るものと考えているからである。検索システムにおけるデータの更新は、システム管理者による複雑な操作と大きな計算機パワーを必要とする、再搭載によって実現する。普通は1年以上、早くも数ヶ月の間隔をおく必要がある。利用者にとっては読出専用 (Read-Only) のものと言える。また、索引語ファイルは原データよりも大きい容量を必要とする。それらの理由で、汎用計算機⁽¹¹⁾でしか利用できない。

歴博もそうであるが、学術データベースの大部分はこのような検索システムによってサービス⁽⁸⁾されている。どうしてこのような状況になったのか、二三の関係者にたづねてみたが、納得のいく答は得られなかった。うまく説明できないが、全体として何か偏っていると思われるのである。例えば、最近急速に発達したワークステーションが、ハードウェアとして能力充分であっても、現行の学術データベースを簡単に移行できるとは思えないのである。

自分なりに先程の疑問に答えておくことにするが、このことについては是非御教示に与りたいと思う。

- (1) 初期に固まったのを簡単に変えられない。
- (2) 新しい、よいソフトが現われれば代えるが、まだ現われていない。
- (3) データ著作権の保護に適している。
- (4) 新しいデータ方式 (テキストデータ、画像データなど) は別システムにより、いま問題にしているような目録データだけは従来通りとする。
- (5) データベース技術の発達が、一般の計算機技術にくらべて遅い。しかも、学術データベースのことまで考えてくれない。

6 データづくり論——2, 3の提言

筆者がデータづくりをはじめた1955年頃はまだパソコンもなく汎用計算機で行うのが当然の時代であった。開館・総合展示の準備の時期に、経費的にもほとんど顧みられなかったことは、その後、数年にわたって推進の障害となった。[3]はその頃の産物であり、本データはこれがやむ好転したときに開始された。いま、あたかも当時からこんにちのような便利なパソコンがあったかのような前提に立って、過去十年余を批判されることがある。全く当を得ないことと言わなければならない。

ところで、日本の計算機ユーザは実力が低いと言われる。筆者の周囲は当然のこととして、理科系の学術研究者、民間企業でさえもそれは例外ではない。外国製のソフトウェア、ハード・ソフトメーカの開発力に頼っているのである。もっとも、パソコンの普及以来、ユーザが自力で開発することが急激に多くなってはいる。歴博でも、少なくともユーザ自身が計算機を動かすことからはじめ、実力を向上してもらう時期に来ているのではないだろうか。このことについて、最近の企画展示フォーラム^[10]で石田晴久氏の、つぎの発言をよく考えてみる必要がある。

「……使いやすいということが挙げられるのですが、これは一昔前にくらべればの話でありまして、こういうものは（やはり）コンピュータですので、だれでも何も考えずに使えるというものではないですね。よくユーザのかたからはコンピュータを研究している者への要望として、パソコンなどというものは、だれでも、ド素人でも使えるようにしてくれ、という話がでるんですけども、これは、また、実際上は無理だと思うんです。とにかく、コンピュータでやる仕事はけっこう複雑ですから、何も勉強——というか、修練もしないで使うのは無理で、ワープロを使った経験のある方はすぐわかると思うのですが、日本語の場合にはどうしても『かな漢字変換』をしなければならぬのですが、その変換のやり方は、やはり、ある程度練習して身につけないと使えない。……」

このような修練は、かな漢字変換からはじまってプログラミングにいたるまで数々必要である。それを自分の研究のまわり道と言うようでは困るのである。

つぎに問題となるのは、データづくりと、つくられたデータの扱いである。説明の都合上後者から先に述べる。

データの公開と言うと、完成されたデータベースの公開と受けとる人が多い。往々にして著作権が公開の障害になり、それへの慎重な対応が必要である。ここで問題とするのは、「完全な」データへのこだわり、および、諸々の理由によるデータの出し渋りである。ここでもう一度、石田氏の発言を引用する。^[10]

「……考えてみますと、図書館と言うのは情報の倉庫ですけども、お金をとりませんね。そういう意味で、データベースは一種の電子図書館であると解釈して、誰でも自由に、無料でアクセスできるような形態にするのが望ましいのではないかと思います。データベースをつくる方からいいますと、折角、手間ひまかけて苦勞してデータベースをつくってコンピュータに入れたわけですから、なるべく多勢の人に使ってもらった方がよろしいわけですね。ですからそういう意味で、日本ではデータベースというと、何か大それたものだという感じがあって、それを使うのは一部の限られた人であって、しかもお金がかかるということだったんですけども、これからはどんどんオープンにしていなければありがたいと思うのです。……」(中略)「……このような博物館の場合にも、博物館の先生方だけですべてやろうというのは、なかなか大変なプロジェクトもあるに違いないので、そのときには全国にいる文化財愛好者に力を合わせてやってもらうという可能性もあるのではないかと思います。……」

データは完成品を公開して利用してもらっただけでなく、改良、増強を広く呼びかけて、多くの人々の共同作業として完成に達するようにはならないのだろうか。技術的に言えば、現在の全国ネットワークは、それを充分、容易にサポートできる状況にある。ただし、先程述べた「実力」をもう少し上げていただく必要があるように思う。これが円滑に行われるようになると、はじめから遠隔地同士で新しいデータを共同でつくることもすぐ可能になるであろう。データづくりは結局お互いのサービスなのではないだろうか。

7 おわりに

東大寺文書目録データは、故・田中稔・前情報資料研究部長の要望からはじまり実現した。なかなか手強い対象と思われたので、国宝重文目録データ^[3]をつくっている間待っていただいた。入力の最中に、いろいろ御教示を仰いだが、その都度、実に楽しそうに説明されたのが印象的であった。一通りデータファイルが動くようになり、早速 ASPEN で検索できるようにコマンドファイルを用意したところ、部長室の端末でよく試して居られたほか来客等にもそれを示して喜んで居られたようである。検索システムへ搭載の細目をつめる打合わせに出席していただいたあとは、時々進行状況を報告するだけになってしまった。試作版が動いたのは亡くなられた後で、誠に申しわけない次第である。

この報告も、追悼号に間に合わず、重ねて申しわけないことになってしまった。遅きに失したが、ここに追悼の言葉とする次第である。

綾村宏氏（奈良文化財研究所 歴史研究室長）には、目録の細部にわたり数々の御教示に与り、また、オンライン公開に当り両機関間の調整の労をとっていただいた。ここに厚く御礼申しあげる次第である。

オンライン公開にいたるまでに必要な諸業務は、つぎのように担当した。

- データファイル
 - 入力 吉岡実*, 丸山昭（株・カンテック）[*当時、現(株)アイビーエス]
 - 校正 吉田悦子, 是洞三栄子, 山路正子ほか8名
 - 処理 照井武彦
- 検索の手引き
 - 執筆 照井武彦
 - 組版 佐藤安一郎, 森谷文子（情報処理係）
- 検索システム搭載
 - 設計 矢木克己, 中野良一（日立公共システムエンジニアリング・株）
 - 処理 出口秀男（同）

註

- (1) [3]の中に入れ子構造の顕著な例を述べ、ひとつの対処の仕方について報告した。
- (2) オブジェクト指向データベースはこれからのよい解として期待しているが、まだこれに触れる機会

に恵まれていない。

- (3) Desk-Top Publishing の訳語。中でも論文によく利用される TEX が研究者の間で普及して居り、学生教育にも最近とりあげるようになった。現在のところ、日本語文用はもとの英文用を少し改造したものにすぎないから、この報告のようなデータには適しない。まだ、縦組みもできていない。
- (4) 各端末の通信ソフトに「JIS 83」とか「新 JIS」とかの設定が必要である。およびホストとの接続開始にあたって ENTER TERMINAL TYPE に対して JIS 83 と応える必要がある。
- (5) これに対して、かなまたはローマ字検索・漢字表示による検索システムもあり、これには当然漢字データに読みをふるが、歴博ではこれを採用しているものは少ない。
- (6) これに対してメニュー方式は親しみやすく感じられる半面、表示されるメニューの前後関係の見通しが悪くなりがちで、同じようなメニューがくりかえし現われ、意図通りに進んでいるのか不安になることが多い。特に1メニュー1画面の場合にこれが著るしく感じられる。
- (7) この処理には、田嶋一夫氏作成の JIS 漢字ソーラスを利用した。記して謝意を表わす。
- (8) 検索システムが誕生した西欧語圏では必要のないことで、検索語は勞せずして得られる。日本語をこれに合わせようとする単語分割が必要になるが、これには多くの勞力と知識が不可欠で何時でも実行できるものではない。最近ソフトによる自動分割が試みられるようになっている。^[5]
- (9) 国文学研究資料館では、試験研究で試作した CDROM データベースでこれを実現した。
- (10) 1992年10月公開のあと2〜3年先を目標にデータの追加・増強を予定している。その主なものは日付コードである。最初は項目04からプログラムによって一括抽出するように考えたが、そのプログラムは極めて実現困難のようで、別の手段を考慮中である。
日付コードの主な目的は、検索集合のソートである。目録第6巻の編年目録を、その間活用していただきたい。
- (11) しかも、どの汎用機でも利用できるわけではない。我が国で可能なのは、知る限りでは、STAIRS (IBM), FAIRS (富士通), ORION (日立), CIR-J (ユニシス) の4つである。その他の機種が入っている大学のセンターではその都度ソフト開発して居られるようである。
- (12) 一方では手作りの1ボードマイクロコンピュータの盛んな時期であったが、いま、議論しているようなデータとは全く結びつかないものであった。他方、日本語ワープロの製品第一号が出現した時期であった。これは500万円以上の価格であった。これはまた、ワープロ同士および計算機とのデータ交換の配慮は全くなかった。これはかなり後まで続いていたように記憶している。

参考文献

- [1] 奈良国立文化財研究所編：東大寺文書目録 全6巻，同朋舎，1979～1984
- [2] 照井武彦：旧高田領取調帳データベースについて，国立歴史民俗博物館研究報告 第12集，1987
- [3] 田辺三郎助 照井武彦 池田宏：国宝重要文化財総合目録（美術工芸品編）のデータファイル作成について，国立歴史民俗博物館研究報告 第16集，1988
- [4] 照井武彦：人文科学研究支援のための日本語・漢字データベースの作成経験～2（書式について），電気関係学会東北支部連合大会，1989
- [5] 研究代表者・小山弘志：国文学データベースの CDROM による出版・利用に関する実用化試験研究，国文学研究資料館，1991
- [6] 国立歴史民俗博物館：データベース利用申請の手引き，1991.4
- [7] 国立歴史民俗博物館：データベース検索の手引き「東大寺文書目録データ」，1992.10
- [8] 全国共同利用大型計算機センターライブラリ・データベース連絡会：オンライン・データベース利用ガイド（第11版）1991.11
- [9] 情報検索システム ORION 解説，同検索の手引，VOS3-AS プログラムプロダクトマニュアル，日立製作所。
- [10] 国立歴史民俗博物館：科学の目でみる文化財，アグネ技術センター，1993，に収録，本稿受付時点で印刷中。

(国立歴史民俗博物館情報資料研究部)

資料1 文書写真の例

『東大寺文書目録』第一巻第1部第1第313号—1「源頼朝御教書」(東大寺図書館掲載許可済)

資料2 目録本体の例
『東大寺文書目録』第一巻 96, 97頁

0005065 1096 101313 000000 01 三一三
 0005066 1096 101313 000000 02 源頼朝御教書等
 0005067 1096 101313 000000 05 三二・九×一九〇・五
 0005068 1096 101313 000000 06 四紙
 0005069 1096 101313 000000 07 平安・鎌倉、
 0005070 1096 101313 000000 08 卷子本、
 0005070 1096 101313 000000 30 連券（二通）、
 *0005071 1096 101313 000000 30 表紙・軸後補、（表紙外題）「文章 \geq Δ 解状趣之書 \blacktriangle ∇ 年預五師給之書 \blacktriangledown \leq 」継目裏二黒円印（
 0005071 1096 101313 000000 30 江戸時代）アリ、
 0005072 1096 101313 010000 01 一
 0005073 1096 101313 010000 02 源頼朝御教書（大江広元奉書）
 0005074 1096 101313 010000 04 \geq 〇（別筆）● \leq 〈寿永三年〉七月二日
 0005075 1096 101313 010000 05 三〇・一×九六・〇
 0005076 1096 101313 010000 06 二紙
 0005077 1096 101313 010000 07 平安院政期、
 0005078 1096 101313 010000 08 竪紙（現装貼継）、
 0005079 1096 101313 010000 10 （書出）「解状之趣令経御覽畢ノ条ノ事ノ一北陸道狼藉事」
 0005080 1096 101313 010000 31 （地名）（脚注）伊賀国鞆田庄
 0005081 1097 101313 010000 11 （書止）「以前条ノ鎌倉殿仰如此仍以執達如件」
 0005082 1097 101313 010000 12 （差出）「散位広元 \geq Δ 奉 \blacktriangle \leq 」 \geq ◇（日下）◆ \leq
 0005083 1097 101313 010000 13 充所ナシ、

資料3 データファイルの例

P0000101313000010010	0010131300
P00001013130000100023	1-1-313(1096)三一三
P00001013130000150018	源頼朝御教書等
P00001013130000250024	32.9×190.5
P00001013130000300010	32.9*190.5
P00001013130000350008	四紙
P00001013130000360014	平安・鎌倉
P00001013130000370010	卷子本
P00001013130000750016	連券(二通)
P00001013130000760098	表紙・軸後補、(表紙外題)「文章 解状趣之書 年預五師給之書 」継目裏ニ黒円印(江戸時代)アリ
P00001013130000800014	伊賀国黒田庄
P00001013130000020006	親
P00001013130000030007	0101313
P03001013130000860001	1
P03001013130000870006	921105
P00001013130100010010	0010131301
P00001013130100100021	1-1-313-1(1096)一
P00001013130100150032	源頼朝御教書(大江広元奉書)
P00001013130100200036	(別筆) <寿永三年> 七月二日
P00001013130100250022	30.1×96.0
P00001013130100300009	30.1*96.0
P00001013130100350008	二紙
P00001013130100360014	平安院政期
P00001013130100370020	豎紙(現装貼継)
P00001013130100450058	(書出)「解状之趣令経御覧畢/条々事/一北陸道狼藉事」
P00001013130100800016	伊賀国韮田庄
P00001013130100500048	(書止)「以前条々鎌倉殿仰如此仍以執達如件」
P00001013130100550034	「散位広元 奉 」 (日下)
P00001013130100600014	充所ナシ、
P00001013130100810014	伊賀国黒田庄
P00001013130100020006	子
P00001013130100030007	0101313
P03001013130100860001	1
P03001013130100870006	921105

- (1) -----
- 親子 :親
 文書番号 / P01 : 1 - 1 - 3 1 3 (1 0 9 6) 三一三
 文書名 / P02 : 源頼朝御教書等
 寸法 / P05 : 3 2 . 9 × 1 9 0 . 5
 紙数 / P06 : 四紙
 時代 / P07 : 平安・鎌倉
 形式 / P08 : 卷子本
 備考 / P30 : 連券 (二通)
 : 表紙・軸後補、(表紙外題)「文章 || 解状趣之書 || 年預五師
 給之書 ||」継目裏二黒円印(江戸時代)アリ
 地名 / P31 : 伊賀国黒田庄
- (2) -----
- 親子 :子
 文書番号 / P01 : 1 - 1 - 3 1 3 - 1 (1 0 9 6) 一
 文書名 / P02 : 源頼朝御教書(大江広元奉書)
 日付 / P04 : || (別筆) || <寿永三年> 七月二日
 寸法 / P05 : 3 0 . 1 × 9 6 . 0
 紙数 / P06 : 二紙
 時代 / P07 : 平安院政期
 形式 / P08 : 竪紙(現装貼継)
 本文の首部 / P10 : (書出)「解状之趣令経御覽畢 / 条々事 / 一北陸道狼藉事」
 本文の末尾 / P11 : (書止)「以前条々鎌倉殿仰如此仍以執達如件」
 差出 / P13 : 「散位広元 || 奉 ||」 || (日下) ||
 充所 / P14 : 充所ナシ、
 地名 / P31 : 伊賀国鞆田庄
 : 伊賀国黒田庄

資料 5 検索出力の例